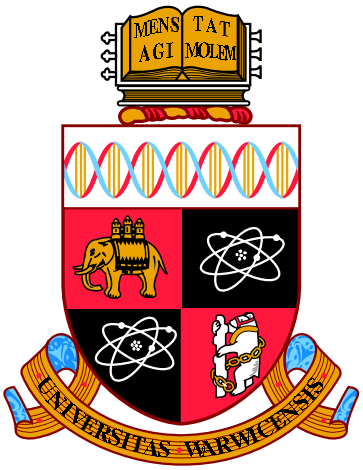


New Mathematical Methods for the Study of Stem Cell Differentiation

by

Elena Camacho Aguilar

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Warwick Mathematics Institute

March 2018

Contents

List of Tables

List of Figures

Acknowledgments

So many people to thank...

First of all, I would like to thank my supervisor for his guidance and support during these years. I very much appreciate the long discussions on such a wide range of mathematical topics. I have learnt so much during these years. I would also like to thank my co-supervisor, Barbel Finkenstaadt, for giving me feedback in my work.

I want to thank the Systems Biology people, in particular Giorgos Minas for his continuous support especially during these last months.

Huge thanks to the friends I have met here, who have supported me and shared very special moments with me, making my days happier. Tianli, Veronika and Mo for being my siblings during the first year in the UK, for the cooking nights and for the many adventures. Thanks to Diana, for the girls nights, the pieces of advice and for sharing the love for cooking with me. Thanks to Gareth for the deep conversations and for being there every time I needed it. To Sara and Costanza for making my lunch time a wonderful break, for all the love, hugs and smiles. To Mim for the delicious breakfasts. To Jake for being such a thoughtful office mate. To Cassa, Paolo, Mattia, Giovanni and Francesco for teaching me the true Italian cuisine, especially the stracciatella. To Steph, Denise and Kate for being such wonderful officemates, for the love, deep conversations and shared rant, and to James for his smiles and cuddles. To Andrea and Aurelija for being the best housemates I could ever had, for the hugs, the laughs, the tears, the deep conversations, the dinners, the movies, etc, we have shared, for feeding me during the last crazy weeks of writing up and for helping me with the move to Houston. To Ale, Elia, Faz, Diego, Felipe, Joe, Clare, Sami, Paul, Eduardo, Simone, Angela, Isty, Daniele...

In addition, thanks to my friends from home who have been always there despite of the distance. To Irene, Isa, María, José Manuel, Manu, Javi and Helena.

A huge thanks to my mentor Francisco Jesús Castro Jimenez, for guiding me all these years not only in the maths but also in life. Also, thanks to José Antonio Langa Rosado for his support and deep converstations every time I go back to the University of Seville. Also thanks to Warmflash Lab for being patient with a mathematician who wants to work in a lab, for teaching me so much during the last two months and for making my new life in Houston fun.

Last but not least, thanks to my family, Elena María, María Elena, Raúl, Lucía for being there and digitally bringing me sushine to UK. To my *American grandparents* Linda and Earl who I love and admire so much, who are always there and with whom I can't wait to share many more adventures. This thesis would not have been possible without the support and guidance of my parents, Manuela and Miguel, and my brother, Miguel, who are the constants in my life. For teaching me values and helping me to be the person I am today, for the energising hugs and for being always there. I cannot express how lucky I am for having their support and understanding even in the most difficult moments.

Declarations

I hereby declare that this thesis is the result of my own work, except where stated otherwise. The thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy and has not been submitted in any other application for any degree.

- Chapter 5 is collaborative work with Dr. James Briscoe and Dr. Robert Blassberg at the Francis Crick Institute.

Elena Camacho Aguilar, March 2018

Abstract

The question of how the fertilized egg develops into an adult organism is one of the most fundamental ones in Biology. A very important stage in the development of the embryo is cell differentiation, in which unspecialised cells, called stem cells, become specialised ones, such as skin or nerve cells depending on the signals that they receive. This is controlled by a very large network of genes that interact with each other, the state of which defines the characteristics of the cell. With the recent development of experimental techniques that allow us to obtain very detailed information about the changes in cells, new data analysis methods and mathematical models are required for the understanding of stem cell differentiation. A common approach to the mathematical modelling of stem cell differentiation is by means of gene regulatory network (GRN) models describing the gene regulation behind the process. However, the number of variables and parameters in these models rapidly scales up as one tries to study more genes in the network, difficulting its analysis. This thesis aims to assess these problems and it is structured into two main parts. In the first one, which comprises Chapters 3 and 4, we will develop a phenotypic quasi-potential landscape model for vulval development in *C. elegans* to illustrate how catastrophe theory can be a powerful tool to construct and understand these recently emerging types of models. Moreover, will use advanced statistical techniques to fit the built model to the experimental data. The second part, in Chapter 5, will be devoted to developing a methodology to understand protein expression data in order to reverse engineer the gene regulatory network from it and create a mathematical model that explains such experimental data.

Abbreviations

ABC	Approximate Bayesian Computation
ABC SMC	Approximate Bayesian Computation Sequential Monte Carlo
AC	Anchor Cell
ANP	Anterior Neural Plate
C. elegans	*Caenorhabditis elegans*
CHIR	CHIRON99021
CME	Chemical Master Equation
CNS	Central Nervous System
DNA	Deoxyribonucleic acid
ESC	Embryonic Stem Cell
FACS	Fluorescence-Activated Cell Sorting
GIGLE	Gene i ordered Gene l Expression
GRN	Gene Regulatory Network
hESC	Human Embryonic Stem Cell
LNA	Linear Noise Approximation
MCMC	Markov Chain Monte Carlo
mESC	Mouse Embryonic Stem Cell
MH	Metropolis-Hastings
MP	Mesodermal Progenitor
mRNA	messenger RNA
NMP	Neuromesodermal Progenitor
NP	Neural Progenitor

TF	Transcription Factor
ODE / ode	Ordinary Differential Equation
RNA	Ribonucleic acid
SDE / sde	Stochastic Differential Equation
VPC	Vulval Precursor Cell
WT	Wild Type

Chapter 1

Introduction

Have you ever wondered how our bodies develop? How from the fertilised egg that sequentially divides into two cells, four, eight cells and so on such a variety of tissues and organs appear?

A key stage in the development of our bodies is cell differentiation, in which unspecialised cells, called stem cells, become specialised ones, such as skin or nerve cells depending on the signals that they receive. This is controlled by a very large network of genes that interact with each other (Davidson 2010), the state of which defines the characteristics of the cell. However, this process is still not completely understood.

In 1957, Waddington, a renowned embryologist, pictured this process in a very intuitive way (Waddington 1957). He imagined a differentiating cell as a marble rolling down a landscape of hills and valleys. The marble would start its journey at the top of the landscape, which would represent the unspecialised state of the cell. As the marble rolls down it encounters some bifurcation points where the path divides into two or three branches. The marble would take one of them and will continue its journey until the bottom of the landscape, where there would be different valleys, one for each type of cell present in the body of the organism. The valley where the marble ends up defines the type of cell that the stem cell differentiated to. The topology of the landscape was controlled by the network of genes which, in his mind, represented tight ropes pulling from the landscape producing these branches or chreodes. This metaphor is refered to as **Waddington's epigenetic landscape**.

Some years later, in 1975 the mathematician René Thom, inspired by Waddington's idea, developed the formalism to translate it into mathematics (Thom 1975). He suggested using the topological theory of dynamical systems to model discontinuous changes in natural phenomena and understood morphogenesis as a structurally stable (i.e. robust or insensitive to small perturbations) type of bifurcations. This theory was further developed and applied by Christopher Zeeman (Zeeman 1976, 1977) giving birth to what is known as **catastrophe theory**. These mathematical ideas generated great interest at the beginning but they were later taken with confusion and controversy, their applications being critisized because

they were "purely qualitative" (Guckenheimer 1978, Poston & Stewart 2014). Unfortunately, these ideas were not applied in Biology very much.

With the recent development of experimental techniques that allow us to obtain very detailed information about the changes in cells, new data analysis methods and mathematical models are required for the understanding of stem cell differentiation.

Attempts have been made to map the gene regulatory network (GRN) controlling the process of stem cell differentiation. However the number of links and the feedback within the system make intuitive understanding difficult to obtain. In order to assess this problem, different computational and mathematical models have been developed. A common approach to the mathematical modelling of stem cell differentiation is by means of reaction networks describing the gene regulation behind the process (Mogilner et al. 2006, Craciun et al. 2006). In these *mechanistic* or *quantitative* types of models the variables represent the state of activity of each gene, which in turn depends on the state of the other genes that interact with it. The stable states of the cells can be regarded as stable states or attractors of the model. One could study how these attractors depend on the environment or signals that the cell is receiving, which are translated into the mathematical parameters of the model. These models have been successfully used to gain insight into how the GRNs control stem cell differentiation and suggest explanations for behaviours otherwise difficult to understand, including self-organisation, oscillations and spatial patterning (Kondo & Miura 2010, Novák & Tyson 2008, Verd et al. 2018). However, these types of models require a large number of parameters and variables which rapidly scales up as one tries to study more genes in the network, complicating its analysis.

With this in mind there has been an increasing interest in new kinds of mathematical models recently that bring back the epigenetic landscape metaphor (Corson & Siggia 2012, 2017, Corson et al. 2017). These models step back from the molecular scale and reason directly on the phenotype expressed in geometrical terms, focusing on the dynamics of the general process rather than on the deep molecular scale. The idea is to define a landscape that contains valleys or basins of attraction for each type of cell that the stem cell can become. The sizes of the valleys, defined by the depths of the attractors, are related to the signals that the cell is receiving which push it to become one type of cell or the other. Although it seems like a very qualitative approach, it certainly allows to not only quantitatively fit experimental data but also provides some predictions about the process (Corson & Siggia 2012, 2017, Corson et al. 2017). These models contain the *essence* of the process that is necessary for its understanding. However, an open question is how

to build these landscapes. These models are characterised by the ways in which the attractors appear and disappear, i.e. the allowed bifurcations, and how the biological signals are mapped into the parameters of the mathematical model; and this brings us back to catastrophe theory.

Catastrophe theory provides very powerful tools for classifying the types of bifurcations or singularities that can be present in a landscape, that in mathematical terms can be expressed as a family of potential functions. In fact, Thom's classification theorem states the types of singularities that a family of functions with less than 5 parameters can have.

The content of the thesis is as follows. In Chapter 2 we will provide the necessary background to follow the arguments presented in the next chapters of the thesis. The first section will be devoted to present the biological problem that we aim to understand, i.e. cell differentiation, to introduce the fundamental biological terminology that will be used throughout the thesis and to summarise some experimental techniques that are currently available to study the process of cell differentiation. We will then move into the mathematical background. First, the classical reaction networks approaches that are used to tackle these problems, both deterministic and stochastic approaches, will be explained. The high dimensionality of these systems will then motivate the introduction of the quasi-potential landscape models. In this regard, we will give a short introduction to catastrophe theory that we will use for the development of a particular model of this type. Finally, the last section will be devoted to describe particular statistical methods used for the estimation of the parameters of our model using the data. Even though the material in this chapter is not necessarily presented in the order in which it will be used in the main chapters, we thought that this was a more natural order.

In Chapter 3 we will motivate the use of catastrophe theory in the development of quasi-potential landscape models by proposing a model for vulval development in *C. elegans*. A biological background for this process and a review of the previous modelling approaches will be given at the beginning of the chapter. Inspired on a previous quasi-potential landscape model proposed by Corson & Siggia (2012) and, by taking advantage of catastrophe theory, we will tailor a new landscape inferred from the observed behaviour of the system. We will motivate and describe the chosen landscape's topology and we will show that this topology can be obtained by combining two elementary catastrophes. We will then show how to link the real system to the mathematical abstraction by mapping the biological signals into the three control parameters of the model. The chapter will finish with a description of the numerical implementation of the model that allows us to simulate

the system under different experimental conditions.

In Chapter 4, we use statistical methods to estimate the parameters of the model proposed in Chapter 3 by using the experimental data available in the literature. We will explain the implementation of the statistical method chosen for these purposes and will present some results obtained from it. We will show that our model can reproduce two subsets of the data available separately and it does so very succesfully. On the other hand, we will show that we are not able to make the model compatible with the two subsets of the data together and we will discuss why we think this happens.

Chapter 5 will be devoted to the study of the gene regulatory network controlling neural and mesoderm specification of mouse embryonic stem cells (mESCs). From the protein expression data povided by our collaborators, Dr. James Briscoe and Dr. Robert Blassberg at the Francis Crick Institute, we will aim to understand the underlying gene regulatory network. In order to make sense of the data, we develop a new method for the visualisation of gene and protein expression data that we call Gene i ordered Gene l Expression (GIGLE). This method will allow us to observe the dynamics of the protein expression of a population of cells in time. With this method, we will show that under specific experimental conditions the biological system is bistable and we will provide a minimal gene regulatory network that reproduces the features observed in the data, this bistability in particular. We will provide a deterministic mathematical model for this gene regulatory network and will show that it can qualitatively reproduce the data under three different experimental conditions. We will also propose a stochastic version of the model and give some results obtained from it. We will finish this chapter by discussing the results and provide some exciting new directions.

Finally, in Chapter 6 we will conclude with a brief summary of the work presented.

Chapter 2

Background

2.1 Biological background

The question of how the fertilized egg develops into an adult organism is one of the most fundamental ones in Biology. It has been observed that once the egg is fertilized, it sequentially divides over and over again generating more and more cells that organise themselves into different structures that constitute tissues, organs, and finally the complete adult organism. But, how does this happen?

In the development of an adult organism, four main processes occur: pattern formation, morphogenesis, cell differentiation and growth. **Pattern formation** is the process in which the body plan is set up; cellular activity is organized in space and time so that a well-ordered structure develops whithin the embryo. During **morphogenesis** the embryo goes through changes in its form that mark the beginning of shape creation. **Cell differentiation** is the developmental process by which cells become functionally and structurally different from each other, giving rise to different types of cells such as muscle, skin or blood cells. Finally, **growth** is the developmental process in which the embryo substantially increases in size. In this thesis we will focus on analysing the third process: cell differentiation.

2.1.1 Cells

Cells are the building blocks of living organisms; plants, insects, animals are all a compound of different types of cells. A cell can be structured into a membrane, a cytoplasm and a nucleus that contains the genetic information encoded into DNA. DNA is a double-stranded helix that contains the information that a cell needs to produce proteins, which perform most of the functions in living organisms. This genetic information is encoded into a sequence of a 4-letter alphabet (A,T,C,G). Each of these letters corresponds to a molecule: adenine (A), thymine (T), cytosine (C) and guanine (G). The subsequence of DNA that is comprised of the right sequence to encode a particular protein is called **gene**.

A protein-coding gene comprises a coding region (the part of the DNA that

contains the instructions for making the particular protein) and adjacent DNA sequences that control the synthesis of such protein. **Gene expression**, the process by which a protein is synthesised, comprises two steps: **transcription** and **translation**. During transcription, a *machinery* called RNA polymerase enzyme binds to the control region at the beginning of the gene, called **promoter**, and copies the coding region of the gene into a single RNA chain called **messenger RNA** (mRNA). This mRNA molecule moves out of the nucleus into the cytoplasm in which translation takes place. During translation, the mRNA is *read* by other *machines* (the ribosomes) which create the protein. This idea that genes are first transcribed into mRNA and then translated into protein (gene $\rightarrow$ mRNA $\rightarrow$ protein) is called the *central dogma* (Watson & Crick 1958), and will be an important notion in this thesis.

Genes interact with each other by affecting each other's expression. As we mentioned before, there are adjacent regions in a gene that control the synthesis of the protein that it codes for. Product proteins of other genes can bind to these control regions to either induce (activation) or repress (inhibition) transcription. Proteins that control the expression of a gene are called **transcription factors**.

The interaction between genes can be described by a network in which the nodes are genes that are connected to each other if they affect their expressions. These networks are often called **gene regulatory networks** (GRN) (Levine & Davidson 2005). For example, a gene A can encode a transcription factor that induces the expression of gene B. This implies that if gene A is expressed, and no other transcription factors that affect the expression of B are present, then B will be expressed. On the other hand, gene C can encode a transcription factor that inhibits or "switches off" gene A. This means that if C is expressed then A will not be expressed, or at least its expression will be reduced. One can do a graphical representation of such network. In this case, gene A and gene B would be connected by a line as well as gene C and gene A (see Figure 2.1). We will discuss more about GRNs later in this section.

Figure 2.1: *Example of regulation between three genes. A induces B. C represses A.*

Regarding the size of the network, every cell in mammals contains $\approx 25,000$ genes in its DNA (Huang & Kauffman 2012, Claverie 2001, Pennisi 2003, Southan 2004). However, not all of the genes interact with each other. In fact, only $5-10\%$ of the genes in the genome[1] encode transcription factors (van Nimwegen 2006, Huang & Kauffman 2012). If we also include parts in the genome that code for microRNAs

[1]Genome is the set of all the genes in an organism

(miRNA), which are transcribed into RNA but do not encode proteins and regulate the expression of genes, the network consists of around 3,000 genes instead of the 25,000 genes of the genome (Huang & Kauffman 2012). Nevertheless, this is still a very big network.

There is some controversy on whether the gene regulatory network is constant in time, i.e. on whether the interactions between the genes in the network change, in an organism, in a life time. In the view of some biologists, the network is not fixed but continuously changes its connections during life time (Lu et al. 2009). While other group holds that the network is fixed unless a mutation happens (Huang 2012).

Cells can also interact with each other by intercellular communication, also known as **cell-cell signalling**. An inducing signal can be transmitted from one cell to another in three main ways. The first way is by a diffusible protein, where the cell sending the signal secretes a diffusible protein to the extracellular space. The receiving cell would have a receptor in the membrane that can bind to that specific signalling molecule. The second way is by direct contact where the cell sending the signal and the cell receiving it have two complementary proteins in the surface that bind together if the cells are sufficiently close. The third type of signalling can be done through gap junctions where the two cells have a *tunnel* connecting each other through which the signalling protein passes. In any of the three ways, once the signal arrives to the receiving cell, it produces a **cascade** of reactions which alters the pattern of gene expression in the responding cell (Wolpert et al. 2015).

The understanding of the mechanisms of gene expression and cell-cell interactions will be useful in the discussion of the original question about cell differentiation.

2.1.2 Cell differentiation

All the somatic cells in an embryo are derived from the fertilized egg or zygote by successive rounds of divisions. Thus, with rare exceptions, they all contain the same genetic information, the same as in the zygote. This is the principle of **genetic equivalence**. The differences between cells are therefore generated by differences in their gene expression which controls the protein synthesis. This is the principle of **differential gene expression**. This was shown by John Gurdon and Shina Yamanaka who were awarded the Nobel Prize in 2012 for their work on stem cells[2].

Cell differentiation is controlled by the underlying gene regulatory network comprising transcription factors and intercellular signals discussed earlier. During cell differentiation, stem cells change their gene expression by dynamically switching

[2]A stem cell is a cell that has the potential to self-renew by cell division, maintaining its differentiation potential, and to create the family of cells of an entire tissue.

on and off different genes. This defines a certain state or identity which is reflected in its molecular organisation. In particular in the proteins that are present in the cell. What function a cell can perform largely depends on the proteins that it mostly produces. For example, the hemoglobin in red blood cells enables them to transport oxygen, or myosin proteins allow skeletal muscle cells to contract. The process by which stem cells become more and more specialized, losing their competence to proliferate and diversify, is called **differentiation**.

The particular type of cell which a cell differentiates to is called **cell fate**. A cell or tissue is **competent** when it can sense signals that direct it towards a particular fate. A cell or tissue is said to be **specified** or **committed** to a fate when the signal can be removed with no change in the fate outcome. Cells are **determined** when signals cannot direct them to other fates.

In multi-cellular organisms, the omnipotent zygote (fertilized egg) differentiates via **totipotent** and **pluripotent embryonic stem cells** and **multipotent tissue stem cells**[3] into different mature cell types of the adult body. This process is achieved by a series of *decisions* in which a stem cell becomes gradually more and more specialized into a particular cell type.

Embryonic stem cells can give rise to hundreds of cell types by a series of binary decisions or branching events in which the cell decides between typically two lineages, committing to one of them. As an analogy, Huang & Kauffman (2012) propose that one can imagine this process as a branching tree in which the fertilized egg is the main trunk and the stem cells are represented at the branching points of the tree. A cell that can branch into various lineages is multipotent. The last branch points represent **progenitor** cells which are no longer stem cells (they cannot infinitively self renew) but they still can chose between a few cell types. For example, we will study neural and mesodermal progenitors in Chapter 5. The very ends of the branches of the tree represent the mature, terminally differentiated cell types of the body, such as red blood or liver cells.

It is important to note that cell types are *discrete* cell identities. Even though the process of cell differentiation corresponds to a continuous change of gene expression, different types of cells are distinct from each other, i.e. they are **phenotypically** different from each other[4]. As Waddington, a renowned embryologist of the last century said in the 1940s: Cell types are "well-recognisable" entities and

[3] Totipotent or omnipotent stem cells have the potential to develop into any type of cell in an organism. Pluripotent stem cells have the potential to develop into almost any type of cells in the organism. Multipotent tissue stem cells are stem cells that can give rise to other cell types but they are limited in their ability to differentiate, being limited to a specific tissue type.

[4] Genotype refers to the genetic endowment of a cell or organism. Phenotype corresponds to the visible appearance, internal structure and biochemical characteristics.

"intermediate states are rare" (Waddington 1957, Huang & Kauffman 2012). Cell types are well separated in the *phenotype space* and are stable. However, the way a cell gradually changes its gene expression in order to become a certain type of cell is not well understood.

2.1.3 Biological techniques for the study of cell differentiation

In order to study cell differentiation biologists observe the embryonic development of a range of model organisms, which can be seen as representatives of the different phylogenies. For example, the fruit fly *Drosophila melanogaster* and the nematode *Caenorhabditis elegans* are the principal model organisms for invertebrates while the frog *Xaenopus laevis*, the mouse *Mus musculus*, the chicken *Gallus gallus* and the zebrafish *Danio rerio* are the main model organisms among vertebrates.

Their development is studied ***in vivo***, i.e. by looking at how the organism develops, or ***in vitro***, i.e. by extracting embryonic stem cells from an organism and observing their development in a petri dish.

In order to study the gene regulatory network and mechanisms that control the differentiation of stem cells, it is possible to create **mutant** organisms or cells for which the DNA has been changed so that, for instance, a certain gene is not expressed. Signals can also be modified by, for example, ablating the source cell, as we will see in Chapter 3.

In addition, there exist various techniques to observe gene expression and protein synthesis in cells. The most versatile and powerful technique for localizing proteins is **fluorescence staining** of cells. With this technique, a fluorescent dye is coupled with an **antibody** that targets a protein of interest. When a fluorescent dye-antibody complex is added to a cell it binds to the protein of interest and emits a fluorescent light of specific wavelength once excited with the correct wavelength. This technique is called **immunofluorescence microscopy** (Alberts et al. 2002). A cell or specimen can be stained with several dyes that fluoresce at different wavelengths, which allows to study multiple proteins.

The levels of fluorescence can then be measured with different techniques, depending on the information that one would like to obtain. For example, **flow cytometry** (see Figure 2.2), is a technique that measures the cellular abundance of a specific protein in an entire population at the resolution of individual cells. It can measure such values for populations of thousands of cells. On the down side, one cannot track the evolution of gene expression for a particular cell in time. We will come back to this technique in Chapter 5. The temporal changes of protein synthesis in time can be measured by using **single-cell real-time imaging**. However, this

technique can only capture up to hundreds of cells at a time, which is normally not enough data to infer properties about the population.

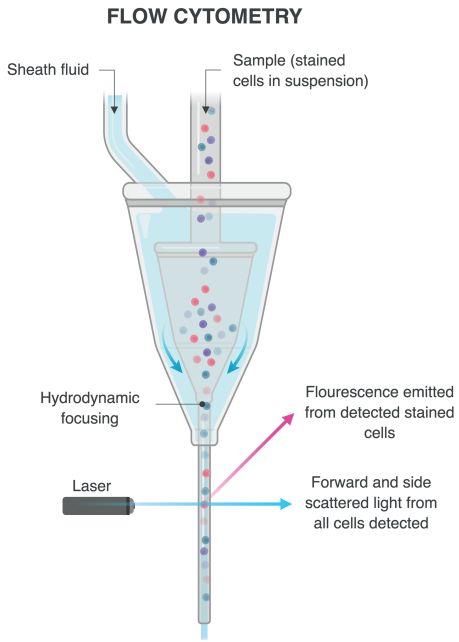


Figure 2.2: *Flow cytometry. A concentrated suspension of labeled cells (by fluorescence staining) is mixed with a buffer (the sheath fluid). Once the sheath fluid is running at a laminar flow, the cells are injected into the center of the stream, at a slightly higher preassure. This way the cells align and pass one by one through the laser beam. Both the fluorescence light emited and the light scattered by each cell are measured. From the measurements of the scattered light, the size and shape of the cell can be determined. From the fluorescence light, one can measure the abundance of fluorescent particles in the cell. Figure made with BioRender.*

In the following sections, we will introduce two very important mathematical techniques that are used to model and study these types of processes. One approach studies cell differentiation by modelling the GRN behind the process as a set of ordinary differential equations. We will introduce these techniques in Section 2.2, and we will put them into practice in Chapter 5. A more recent approach uses the original idea of **epigenetic landscape** given by Waddington (1957) in which differentiating cells are represented by balls rolling down branching valleys or *chreods* in a hill. We will show in Section 2.3 how this can be mathematically expressed. We will give an introduction to catastrophe theory which will prove to be useful for these modelling purposes in Chapters 3 and 4. In the last section of this chapter we will talk about statistical methods for parameter fitting which will be used in Chapter 4.

2.2 Reaction networks

In this section we introduce the reader to reaction networks and their mathematical formulation. Reaction networks constitute a very important concept in Biology and Chemistry since they can represent the interactions between molecules in a biochemical system. They are commonly used in the modelling of GRNs, where they model the evolution of the gene expression in time (Ribeiro et al. 2006, Huang et al. 2007, Perez-Carrasco et al. 2017). Using reaction networks one can study the dynamical behaviour arising from the network.

We first focus on the deterministic approach in which the reaction networks are described by ordinary differential equations (ODEs). In order to motivate this mathematical approach, we study three examples: gene expression, gene regulation and gene regulatory networks. Using the last example, we will describe the concept of stability in dynamical systems.

We will then give a mathematical description for stochastic reaction networks and provide two models for describing their stochastic dynamics.

2.2.1 Deterministic models. Dynamics and stability.

2.2.1.1 Example I: Linear gene expression

Gene expression, as mentioned in Section 2.1, is one of the most important processes in biology. The so-called *central dogma* explains how DNA directs the synthesis of proteins. A gene encoded in our DNA is transcribed to form mRNA molecules in the cell nucleus. The mRNA molecules then travel out of the nucleus of the cell into the cytosol where they are translared into protein molecules. A simplification of the protein production can be described by a reaction system of four processes:

$$\begin{aligned} \varnothing &\xrightarrow{k_1} \text{mRNA} && \text{(transcription);} \\ \text{mRNA} &\xrightarrow{k_2} \varnothing && \text{(mRNA degradation);} \\ \text{mRNA} &\xrightarrow{k_3} \text{protein} && \text{(translation);} \\ \text{protein} &\xrightarrow{k_4} \varnothing && \text{(protein degradation).} \end{aligned} \tag{2.1}$$

Here, we have assumed that the amount of DNA is unchanged and that the molecules required for translation are in excess, so these molecules are not modelled explicitly. The processes with the $\varnothing$ symbol stand for a reaction in which a molecule has been created from (or degraded into) *nothing*, meaning that the origi-

nal (or product) molecule is not modelled in the representation. For example, in the first process, it corresponds to the individual molecules that assemble an mRNA macromolecule. The **reaction rates constants**, k_i, describe the speed of each process.

Let us denote by [mRNA] and [protein] the concentration by unit volume of mRNA and protein respectively. The classical law of mass-action kinetics suggests that the rates of reactions are directly proportional to the concentration of each reactant. For example, in the above reaction network, transcription has a rate k_1 while the mRNA degradation reaction has rate k_2. Assuming that concentrations continuously change over time, the evolution of concentrations is described by the system of ordinary differential equations (ODE)

$$\frac{d[\text{protein}]}{dt} = k_3[\text{mRNA}] - k_4[\text{protein}], \tag{2.2}$$

$$\frac{d[\text{mRNA}]}{dt} = k_1 - k_2[\text{mRNA}]. \tag{2.3}$$

As we can see, the ODEs in Equations 2.2 and 2.3 describing the deterministic form of the reaction system are linear. This is always the case if the reaction involves at most one reacting molecule. Any reaction system that has only zeroth-order (a molecule is created from *nothing*) and first-order (a single type of molecule is transformed into some other molecule(s)) reactions, will be linear. This makes the analysis of the system easier for various reasons including that one can compute the equilibrium points of the system given by Equations 2.2 and 2.3 very easily. However, more realistic biological phenomena cannot be formulated this way, as we will see in the following subsection.

2.2.1.2 Example II: Gene regulation

Let us now study a reaction network that describes the induction of a gene G by a transcription factor T. Normally, a single molecule of T will not be enough to induce the transcription of G. Let us assume that there need to be n molecules of T bound to a macromolecule promoter P of G in order for transcription to start. Let us call nTP the concentration of complexes formed by P and n molecules of T, $n \geq 1$. To describe the binding process, we need to describe the binding of n molecules of T to P. The promoter P can be either bound to n molecules of T, described by the complex nTP, or unbound, denoted by P_0 (for simplicity, we neglect intermediate states were less than n molecules are bound). The total concentration of bound and

unbound P molecules is denoted by P_t, and the conservation law is

$$[nTP] + P_0 = P_t, \tag{2.4}$$

where $[nTP]$ is the concentration of nTP.

The production and degradation of the complex nTP can be described by the following reaction system:

$$\begin{aligned} P + \underbrace{T + T \ldots T}_{n} &\xrightarrow{k_1} nTP; \\ nTP &\xrightarrow{k_2} P + \underbrace{T + T \ldots T}_{n}. \end{aligned} \tag{2.5}$$

The complex nTP is formed by collisions of P with n molecules of T. Hence, the rate of the molecular collisions needed to form the complex is given by the concentration of free P, P_0, and the concentration of T to the power of n:

$$\text{collision rate} = k_1 P_0 T^n. \tag{2.6}$$

The complex nTP, on the other hand, dissociates with rate k_2:

$$\text{dissociation rate} = k_2[nTP]. \tag{2.7}$$

The total rate of change of the concentration of the complex is therefore the difference between the collision and dissociation rates, similar to how we computed the rate of change of protein molecules in the previous section. That is,

$$\frac{d[nTP]}{dt} = k_1 P_0 T^n - k_2[nTP]. \tag{2.8}$$

Now, imagine that the mRNA expression of a gene G is produced by these *activated* promoters $[nTP]$ such that

$$\begin{aligned} nTP &\xrightarrow{k_3} \text{mRNA}; \\ \text{mRNA} &\xrightarrow{k_4} \varnothing. \end{aligned} \tag{2.9}$$

We can then represent the change in mRNA concentration, [mRNA] as:

$$\frac{d[\text{mRNA}]}{dt} = k_3[nTP] - k_4[\text{mRNA}]. \tag{2.10}$$

Now, since the time required for a transcription factor to bind to a promoter is

much smaller than the time needed to transcribe a gene (Alon 2006), we can make a steady-state approximation in which $d[nTP]/dt = 0$ while G is being transcribed. This means that:

$$k_2[nTP] = k_1 P_0 T^n. \tag{2.11}$$

We can now use the conservation equation 2.4 to replace P_0 with $P_t - [nTP]$, hence

$$\frac{k_2}{k_1}[nTP] = (P_t - [nTP])T^n. \tag{2.12}$$

Finally, we can solve for the fraction of bound P, to find what it is known as the **Hill equation**:

$$\frac{[nTP]}{P_t} = \frac{T^n}{K^n + T^n}, \tag{2.13}$$

where we have defined the constant K such that $K^n = k_2/k_1$. The parameter n is known as the **Hill coefficient**. If $n = 1$, we obtain what is called **Michaelis-Menten equation**. We can observe that the Hill and Michaelis-Menten equations reach half-maximal concentration when $T = K$. The function in 2.13 is an increasing sigmoidal in T, and the curve is steeper for higher values of the Hill coefficient n.

We can use this approximation to substitute $[nTP]$ by its approximation in Equation 2.14, obtaining:

$$\frac{d[\text{mRNA}]}{dt} = \tilde{k}_3 f(T) - k_4[\text{mRNA}], \tag{2.14}$$

where $\tilde{k}_3 = k_3 P_t$ is a new rate depending on the total concentration of P molecules, and $f(T) = T^n/(K^n + T^n)$ is a hill function of T.

Hill equations are very often used to model the interactions between the genes in a GRN. Similarly to Equation 2.13, the Hill equation modelling inhibition of gene G from protein I is

$$f(I) = \frac{K^n}{K^n + I^n}, \tag{2.15}$$

which is now decreasing in I.

Let us describe now how to apply this knowledge in a gene regulatory network.

2.2.1.3 Example III: Gene Regulatory Networks. Dynamics and stability.

Now that we have introduced the building blocks for describing the regulation between genes we can tackle the modelling of a gene regulatory network.

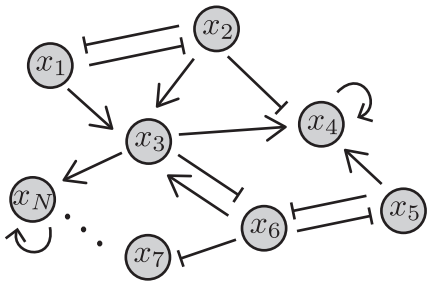


Figure 2.3: *Example of a gene regulatory network with N genes that regulate each other. Triangular arrows represent induction of the downstream node. Flat arrows represent repression of the downstream node.*

The system is comprised by a set of N genes that can regulate each other (see Figure 2.3). The state of the system at time t is represented by a vector $\boldsymbol{x}(t) = (x_1(t), x_2(t), \ldots, x_N(t))$, in a **state space**, where each of the components represent the mRNA or protein concentration values for each of the N genes in the network. The model that describes the change of the state of the system in time is given by the ODE:

$$\frac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}, \boldsymbol{\theta}). \tag{2.16}$$

where $\boldsymbol{F}$ is a function that describes the interactions between the different genes.

For example, suppose that we would like to model the interactions between gene 1 and 2 in Figure 2.3, without taking into account the rest of genes. In this case, the network is formed by two mutually repressing genes ($N = 2$), the state of the system at time t will be a two-dimensional vector $\boldsymbol{x}(t) = (x_1(t), x_2(t))$ and deterministic model would be given by

$$\begin{array}{lll} \frac{dx_1}{dt} & = f_1(x_1, x_2, \boldsymbol{\theta}) & = k_1 \frac{K_2^n}{K_2^n + x_2^n} - k_2 x_1 \\ \frac{dx_2}{dt} & = f_2(x_1, x_2, \boldsymbol{\theta}) & = k_3 \frac{K_1^n}{K_1^n + x_1^n} - k_4 x_2. \end{array} \tag{2.17}$$

Here $\boldsymbol{F} = (f_1(x_1, x_2), f_2(x_1, x_2,), \boldsymbol{\theta})$, where $\boldsymbol{\theta} = (k_1, k_2, k_3, k_4, n, K_1, K_2)$ are the parameters controlling the mutual repression and degradations.

With this model one can investigate which gene expression patterns are most stable and most unstable in the system, i.e. the states for which $\boldsymbol{F} = 0$. This points are called **critical points** of the dynamical system given in Equation 2.16. If the system is in any of these gene expression critical values, it will not change.

In order to understand the type of **stability** of such critical points, i.e. how the system evolves close to these points, one needs to look at the Jacobian matrix of $\boldsymbol{F}$, this is $D_{\boldsymbol{x}}\boldsymbol{F}$, evaluated at the corresponding point. To see this, consider a critical point $\boldsymbol{x}^* \in \mathbb{R}^N$. Let the solution $\boldsymbol{x}(t)$ be near the critical point, i.e. $\boldsymbol{x}(t) = \boldsymbol{x}^* + \bar{\boldsymbol{x}}$

with $\bar{\boldsymbol{x}}$ small. Then, one can Taylor-expand $\boldsymbol{F}$ around the fixed point:

$$\boldsymbol{F}(\boldsymbol{x}(t)) = \boldsymbol{F}(\boldsymbol{x}^*) + D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}^*)(\boldsymbol{x}(t) - \boldsymbol{x}^*) + \cdots,$$

where one can neglet quadratic and higher order terms since we are considering that the solution is close to the critical point, therefore $\bar{\boldsymbol{x}}$ has small norm. Since $\boldsymbol{F}(\boldsymbol{x}^*) = 0$, and $\bar{\boldsymbol{x}} = \boldsymbol{x}(t) - \boldsymbol{x}^*$, one can describe how $\bar{\boldsymbol{x}}$, the change of perturbation from the critical point, changes in time as

$$\frac{d\bar{\boldsymbol{x}}}{dt} = D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}^*)\bar{\boldsymbol{x}}.$$

The matrix $D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}^*)$ is characterised by eigenvalues λ_i, $i = 1, \ldots, N$, that are roots of the characteristic equation

$$\det\left(D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}^*) - \boldsymbol{\lambda}\boldsymbol{I}\right) = 0,$$

where $\boldsymbol{I}$ is the identity matrix. The N eigenvalues are associated with N linearly independent eigenvectors $\boldsymbol{e}_i$. These eigenvectors satisfy that

$$D_{\boldsymbol{x}}\boldsymbol{F}(\boldsymbol{x}^*)\boldsymbol{e}_i = \lambda_i \boldsymbol{e}_i, \quad i = 1, \ldots, N.$$

Therefore, the increase or decrease of the perturbation from the critical point $\bar{\boldsymbol{x}}$ in each direction $\boldsymbol{e}_i$ is determined by the real part of the eigenvalue λ_i.

If the rank of the Jacobian is less than N, then the point is called **degenerate**, otherwise it is **non-degenerate**.

In the case that the point is non-degenerate, the stability of the point can be studied by looking at the eigenvalues of the Jacobian. If any of the eigenvalues have positive real part, the point is **unstable**. In this case, if the system starts close to the critical value, it will move away from it. On the other hand, if the eigenvalues have negative real part, the point is **stable** (an **attractor**), and the system will tend to this value. In the case that the real part of the eigenvalues have mixed signs, the point is called a **saddle** since it has stable and unstable directions. And finally, if the eigenvalues are purely imaginary, the point will be a **center** which will give rise to oscillating patterns of gene expression.

Note that the critical points and their stability will depend on the parameter values. Indeed, it is important to study how the critical points and their stability change as the parameter values change, since it will give information about how the gene expression patterns depend on the structure of the network.

For example, the system in Equation 2.17 is bistable, i.e. it has two stable

critical points, for some parameter values (see Figure 2.4). The system has two stable critical points when the reaction rate constants are all equal to 1, the Hill coefficients are equal to 2 and the half-maximal concentrations are equal to 0.1, i.e. $\boldsymbol{\theta} = (1, 1, 1, 1, 2, 0.1, 0.1)$. We can see that the gene expression pattern towards which the system evolves depend on which gene has a higher concentration at the beginning, as we could intuitively imagine. The symmetry in the flow is given by the structure of the network, i.e. the parameter values.

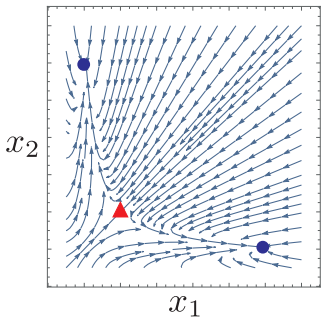


Figure 2.4: *Vector field plot of the GRN with two mutually repressing genes given in Equation 2.17. Arrows represent the vector field determined by the flow $\boldsymbol{F}$. Stable critical points, or attractors, are represented as blue circles. The saddle is represented by a red triangle.*

If, instead, gene 1 repression to gene 2 is stronger than the one from gene 2 to gene 1, the symmetry of the model is broken, and if that repression is strong enough, the system will no longer be bistable and the only critical point of the system will be given by a gene expression pattern with x_1 high and x_2 low.

We will see more examples of stability analysis of a gene regulatory network in Chapter 5.

So far we have just considered deterministic models for reaction networks. However, they are usually stochastic processes. In the following subection we will describe how to derive a stochastic model for them.

2.2.2 Stochastic formulation

Let us consider a well-stirred mixture of $n \geq 1$ molecular species $\{S_1, \ldots, S_n\}$ that interact inside some fixed volume Ω and at a constant temperature. These species can interact through $m \geq 1$ reactions $\{R_1, \ldots, R_m\}$ (e.g. transcription, translation, degradation). Let us call Y_i the number of molecules of the molecular species S_i. The system at time t is defined by $\boldsymbol{Y}(t) = (Y_1(t), \ldots, Y_n(t))$, where $Y_i(t)$ is the number of S_i molecules ($i = 1, \ldots, n$), in the system at time t. Then the j-th reaction can

be written as the process

$$\nu_{j,1}^1 Y_1 + \cdots + \nu_{j,n}^1 Y_n \xrightarrow{c_j} \nu_{j,1}^2 Y_1 + \cdots + \nu_{j,n}^2 Y_n, \tag{2.18}$$

during which the number of Y_i molecules changes from $\nu_{j,i}^1$ to $\nu_{j,i}^2$, $(i = 1, \ldots, n)$. The molecular populations $\boldsymbol{Y}_i(t)$ are now *random variables*. Our goal is to describe the evolution of $\boldsymbol{Y}(t)$ from some given initial state $\boldsymbol{Y}(0) = \boldsymbol{Y}_0$.

If the system is always *well mixed*, it can be shown that there exists a well defined function ω_j for each reaction R_j (Gillespie 2000), called **propensity function** for R_j, such that $\omega_j(\boldsymbol{Y})\delta t$ is the probability (given $\boldsymbol{Y}(t) = \boldsymbol{Y}$) that one R_j reaction will occur somewhere inside Ω in the next infinitesimal time $[t, t + \delta t)$, $(j = 1, \ldots, m)$. The propensity functions ω_j can be written as $\omega_j(\boldsymbol{Y}) = c_j h_j(\boldsymbol{Y})$, where c_j is the **probability rate constant** of reaction R_j specified in Equation 2.18, and $h_j(\boldsymbol{Y})$ is defined to be the number of distinct combinations of R_j reactant molecules available in state $\boldsymbol{Y}$. For example, if the reactions are given by $A \xrightarrow{c_1} \varnothing$ and $B \xrightarrow{c_2} A + B$, then the propensity functions for the reactions would be $\omega_1 = c_1 A$ and $\omega_2 = c_2 B$.

If one R_j reaction occurs then, as we previously mentioned, $\nu_{j,i} = \nu_{j,i}^2 - \nu_{j,i}^1$ new S_i molecules will be produced. The vectors $\boldsymbol{\nu}_j \in \mathbb{R}^n$ are called **stoichiometric vectors**, and contain the total number of molecules of each species produced or consumed in reaction j. This means that if reaction R_j happens in the interval of time $[t, t + \delta t)$, then the system will change from $\boldsymbol{Y}(t)$ to $Y(t + \delta t) = \boldsymbol{Y}(t) + \boldsymbol{\nu}_j$. For example, as before, if the reactions are given by $A \xrightarrow{c_1} \varnothing$ and $B \xrightarrow{c_2} A + B$, then the stoichiometry vectors are given by $\boldsymbol{\nu}_1 = (-1, 0)$, $\boldsymbol{\nu}_2 = (1, 0)$.

Using these quantities, the time evolution of the system can be formulated as a continuous-time inhomogeneous Poisson process with Kolmogorov's forward equation, known in this context as the Chemical Master Equation (CME):

$$\frac{dP(\boldsymbol{Y})}{dt} = \sum_{j=1}^{m} P(\boldsymbol{Y} - \boldsymbol{\nu}_j)\omega_j(\boldsymbol{Y} - \boldsymbol{\nu}_j) - P(\boldsymbol{Y})\omega_j(\boldsymbol{Y}). \tag{2.19}$$

The term $P(\boldsymbol{Y} - \boldsymbol{\nu}_j)\omega_j(\boldsymbol{Y} - \boldsymbol{\nu}_j)$ corresponds to the event that the system is in a different state and, through reaction j, arrives to $\boldsymbol{Y}$. While the term $P(\boldsymbol{Y})\omega_j(\boldsymbol{Y})$ is the probability of the system moving away from state $\boldsymbol{Y}$ through reaction j. Therefore, the state of the system is completely characterised by a set of reaction propensities $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_m)$ and a **stoichiometry matrix** collecting the stoichiometry vectors $\boldsymbol{S} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_n)$, which is an $n \times m$ matrix .

So far, we have considered $Y(t)$ as the number of molecules, but one could be

interested in the species concentration $\boldsymbol{X}(t) = \boldsymbol{Y}(t)/\Omega$. This connection is derived through the system size, Ω, that might be considered to be of the same order of magnitude as the total population size while in chemical systems a natural choice is to use molar concentrations and therefore Ω is regarded as Avogadro's number in the appropriate molar units (e.g. nM^{-1}) multiplied by the volume of the reacting solution in appropriate units (e.g. in litres (L)). As we are going to show, the system size allows us to derive the macroscopic (deterministic) representation discussed earlier as an approximation of the stochastic model. From Equation 2.19, one can derive that the state of the system at some time t is

$$\boldsymbol{Y}(t) = \boldsymbol{Y}(0) + \sum_{j=1}^{m} \boldsymbol{\nu}_j Z_j \left(\int_0^t \omega_j(\boldsymbol{Y}(s)) ds \right) \tag{2.20}$$

where Z_j are independent unit Poisson processes corresponding to the j-th reaction R_j (Anderson & Kurtz 2011, Minas & Rand 2017). Writing this equation in terms of $\boldsymbol{X} = \boldsymbol{Y}/\Omega$, we obtain that

$$\boldsymbol{X}(t) = \boldsymbol{X}(0) + \sum_{j=1}^{m} \boldsymbol{\nu}_j \Omega^{-1} Z_j \left(\int_0^t \Omega u_j(\boldsymbol{X}(s)) ds \right),$$

where $\omega_j(\boldsymbol{Y}) = \Omega u_j(\boldsymbol{Y}/\Omega)$. Using the law of large numbers[5], as $\Omega \to \infty$, we have that $\Omega^{-1} Z_j(\Omega\lambda) \to \lambda$ and thus

$$\boldsymbol{x}(t) = \boldsymbol{x}(0) + \sum_{j=1}^{m} \boldsymbol{\nu}_j \int_0^t u_j(\boldsymbol{x}(s)) ds,$$

where $\boldsymbol{x}(t)$ is the limit of $\boldsymbol{X}(t)$ for $\Omega \to \infty$ (convergence in probability). In turn, $\boldsymbol{x}(t)$ satisfies the ODE

$$\frac{d\boldsymbol{x}(t)}{dt} = \sum_{j=1}^{m} \boldsymbol{\nu}_j u_j(\boldsymbol{x}). \tag{2.21}$$

This approach allow us to move from the deterministic setting to the stochastic setting also. In many cases there already exist existing deterministic models for a given system, with all parameters known in terms of concentrations $\boldsymbol{X}$. If the system size is appropriately derived, one can write the number of species as $\boldsymbol{Y} = \Omega\boldsymbol{X}$, and the propensity functions are derived from the deterministic reaction rates. For example, consider Equation 2.14, which describes the translation and degradation of mRNA as a deterministic system. In order to derive the stochastic formulation,

[5]See Theorem 1.2. in (Anderson & Kurtz 2011)

one could write the number of species as $\boldsymbol{Y} = \Omega \times ([\text{mRNA}], T)$, and use the relation $\omega_j(\boldsymbol{Y}) = \Omega u_j(\boldsymbol{Y}/\Omega)$. The propensity function ω_1 corresponding to the translation reaction is written as $\omega_1 = \tilde{k}_3 \Omega Y_2^n / ((K \times \Omega)^n + Y_2^n)$, and the one corresponding to degradation as $\omega_2 = \tilde{k}_4 Y_1$. This is done based on the fact that the unit of k is always of 1/seconds.

The next step is to find the solution of the CME to describe the distribution of the system in a given time. However, there are very few occasions in which this can be done analytically. A wide number of approaches exist to simulate the system and predict its behaviour. They differ on their exactness and efficacy. We will focus on the Stochastic Simulation Algorithm (SSA) (Gillespie 1977), also known as Gillespie's algorithm, which provides an exact simulation of the system, and on a stochastic differential equation (SDE) approximation that will be used in Chapter 5. For more details on exact and approximate simulation methods we refer the reader to Wilkinson (2011).

2.2.3 Gillespie's Stochastic Simulation algorithm

Gillespie (1977) develops an algorithm for direct simulation of processes described by the CME that takes into account the probabilistic nature of the process by simulating every *jump* or reaction taking place in the system. SSA is given in Algorithm 1.

As we saw in equation 2.20, a sample path of the CME is a Poisson process so the time between successive reactions will be given by an exponential distribution. Given the number of steps N, the initial time $t(0) = t_0$ and the initial number of molecules $\boldsymbol{Y}(0) = \boldsymbol{Y}_0$, the Gillespie algorithm computes a chain of states $\{\boldsymbol{Y}_i\}_{i=0}^N$ of the system by sequentially simulating the reaction that takes place in a certain period of time. In each step, the algorithm samples a uniformly distributed random number (u_1) and an exponential distributed random number (τ) to determine which reaction occurs and the waiting time for the next event, respectively. By properties of the exponential distribution, the probability of a reaction to be chosen is equal to the ratio of its propensity (ω_j) and the sum of all reaction propensities (ω_T). The time of the next reaction, τ, is the minimum of the times of the next reaction of any of the types and it has, by properties of the exponential distribution, an exponential distribution with rate equal to the sum of all propensities.

Although the solution given by Algorithm 1 is an exact simulation of the system, the computational time required is very high. This is the reason why we will use an approximation instead, which is given in the following subsection.

Algorithm 1: Stochastic Simulation algorithm (SSA)

Initialise system by setting $t = 0$ and $\boldsymbol{Y}(0) = \boldsymbol{Y}_0$
for $i = 0, \ldots, N$ **do**
  Compute reaction propensities $\omega_j = \omega_j(\boldsymbol{Y}_i)$ and total propensity $\omega_T = \sum_{j=1}^{m} \omega_j(\boldsymbol{Y}_i)$
  Generate random numbers $u_1, u_2 \sim U(0,1)$.
  Choose reaction r to occur which satisfies $\sum_{j=1}^{r-1} \omega_j < u_1 \omega_T < \sum_{j=1}^{r} \omega_j$.
  Compute $\tau = (1/\omega_T) log(1/u_2)$.
  Set $\boldsymbol{Y}_{i+1} = \boldsymbol{Y}_i + \boldsymbol{\nu}_r$.
  Set $t_{i+1} = t_i + \tau$.
end

2.2.4 Chemical Langevin equation: an SDE approximation.

The chemical Langevin equation approximates the solution of the CME by a stochastic differential equation (Gillespie 2000, Wilkinson 2011, Anderson & Kurtz 2011).

Consider the stochastic process formulation introduced in Subsection 2.2.2. If we assume that there is a small enough δt such that the propensity functions are practically constant in the interval $[t, t+\delta t]$, i.e. $\omega_j(\boldsymbol{Y}(t)) \simeq \omega_j(\boldsymbol{Y}(t'))$ for all $t' \in [t, t+\delta t]$ and all $j \in \{1, \ldots, m\}$, then one can assume that the number of times that a reaction R_j occurs in the time interval $[t, t+\delta t]$ is a statistically independent Poisson random variable $\mathcal{P}_j(\omega_j(\boldsymbol{Y}(t)), \delta t)$[6]. This allows us to write the state of the system at time $t + \delta t$ as:

$$\boldsymbol{Y}(t+\delta t) = \boldsymbol{Y}(t) + \sum_{j=1}^{m} \boldsymbol{\nu}_j \mathcal{P}_j(\omega_j(\boldsymbol{Y}(t)), \delta t). \tag{2.22}$$

That is, the state of the system at time $t + \delta t$ is the result of summing the number of molecules of each species that have changed during the length of time δt to the state of the system at time t (Gillespie 2000).

If one also imposes that δt is large enough such that the expected number of times that each reaction R_j occurs in the time interval $[t, t+\delta t]$ is much larger than 1, then one can approximate each Poisson random variable $\mathcal{P}_j(\omega_j(\boldsymbol{Y}(t)), \delta t)$ by a normal random variable with the same mean and variance. Therefore, taking

[6]The mean of a Poisson counting process is equal to its variance, $\mathbb{E}[\mathcal{P}(\lambda, t)] = \mathrm{Var}[\mathcal{P}(\lambda, t)] = \lambda t$.

this into account, one can rewrite Equation 2.22 as

$$\boldsymbol{Y}(t+\delta t) = \boldsymbol{Y}(t) + \sum_{j=1}^{m} \boldsymbol{\nu}_j \mathcal{N}_j(\omega_j(\boldsymbol{Y}(t))\delta t, \omega_j(\boldsymbol{Y}(t))\delta t), \tag{2.23}$$

where $\mathcal{N}(\mu, \sigma^2)$ denotes a normal random variable with mean μ and variance σ^2. Note that $\boldsymbol{Y}(t+\delta t)$ is no longer an integer random variable, but a real random variable since we have changed the independent integer Poissons $\mathcal{P}_j$ by independent real normals $\mathcal{N}_j$.

Finally, one can rewrite Equation 2.23 as the SDE

$$\mathrm{d}\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{\omega}(\boldsymbol{Y}_t) + \sqrt{\boldsymbol{S}\mathrm{diag}\left\{\boldsymbol{\omega}(\boldsymbol{Y}_t)\right\}\boldsymbol{S}^t}\mathrm{d}\boldsymbol{W}_t, \tag{2.24}$$

where $\boldsymbol{S}^t$ denotes the transpose of $\boldsymbol{S}$ and $\boldsymbol{W}_t$ is an n-dimensional Wiener process or Brownian motion.

Setting δt to a fixed, small value, one can use this approximation to simulate the system with **Euler-Maruyama method** (see Appendix A), whith accuracy and efficiency depending on the choice of such time step. It is also important to notice that since the noise in 2.24 is Gaussian, the simulation can give negative values to $\boldsymbol{Y}_t$. Among the different strategies to address steps of the algorithm that would result in a species' amount being negative, we will apply the following one: set the species count to zero in that case. This approximation will be used in Chapter 5 for the stochastic modelling of the gene regulatory network proposed.

2.3 Waddington's landscape and catastrophe theory

As we saw in the last section, modelling the GRN behind a process and understanding its dynamics can be a difficult task, since it normally comprises a very high dimensional system that depends in a large number of parameters. This is why, recently, a new modelling idea is becoming more and more popular: the quasi-potential landscape model (Huang & Kauffman 2012, Huang 2012, Corson & Siggia 2012, 2017, Corson et al. 2017).

2.3.1 Waddington's landscape as a mathematical description of cell differentiation

The idea dates back to 1957, when Waddington, a rewnowned embryologist introduced the idea of **epigenetic landscape** to describe the developmental pathway of a stem cell (Waddington 1957). In his analogy, differentiating cells are marbles

rolling through a landscape of hills and valleys drawn towards lower points of minimum elevation (see Figure 2.5). A cell starts the journey in a valley at the top of the landscape and, as it progresses downward, it will face intermediate branching valleys that represent the series of fate choices made by the developing cell. Starting at the top valley, representing the omnipotent state, the ball rolls down to intermediate branching valleys which represent intermediate or more differentiated states, finishing at the bottom valleys which correspond to the final differentiated cell type. The topology of his landscape depends on the interactions between the genes, imagined as tight ropes which control the shape of the chreods (see Figure 2.6).

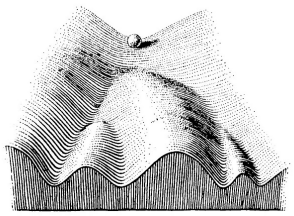


Figure 2.5: *Waddington's epigenetic landscape. Adapted from Waddington (1957)*

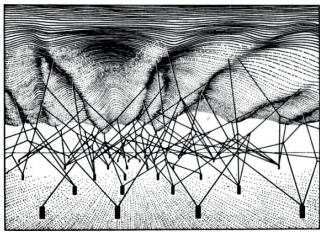


Figure 2.6: *The topology of Waddington's epigenetic landscape is underpinned by the influence of the genes. Adapted from Waddington (1957)*

The quasi-potential landscape is the translation of this idea to dynamical systems concepts[7]. Corson & Siggia (2012) explain it in the following way. The potential function will qualitatively describe how a stem cell changes its phenotype to become a particular type of cell. The attractors, in this case, will correspond to phenotypes or types of cells that the stem cell can differentiate into. We could imagine that the "decision" of the cell is symbolised by the trajectory of a ball on a landscape. This landscape will have valleys or attractors that will represent the possible fates that a stem cell can take at a particular time. The state of commitment

[7]We are aware that there exist various interpretations of the quasi-potential landscapes in biology, some focusing in evolution (Huang 2012, Jaeger & Monk 2014, 2015), other focusing on the gene regulatory network level (Huang et al. 2007, Huang & Kauffman 2012, Zhou et al. 2012) and others on the phenotypic level (Corson & Siggia 2012, 2017, Corson et al. 2017). In this thesis we focus on the latter one but we will discuss more about the differences in Chapter 6.

of a cell is defined by its relative distance to the existing attractors. This landscape will also be dynamic, i.e. only certain attractors will be accessible at particular times of the development. Similarly to Waddington's epigenetic landscape, where the genes control the shape of the landscape, the basins of attraction will be here controlled by the signals to which the cell responds during competence, shifting the boundaries between attractors, biasing the trajectory towards a certain fate. The process of differentiation of a stem cell will then be represented by a trajectory on this dynamic landscape. The attractor in which basin of attraction the trajectory ends, will define the final fate of the cell.

In order to imagine the process, we could think of a ball on a sink full of water. Although not a rigorus simile since water is not a potential system, we think it is a good visual example. Imagine that we were to make a small hole in the base of the sink. The water would begin to drain out of it and the ball would move towards the hole. We will call this drain an **attractor**. We could also imagine a sink with two drains. In this case, if the ball was in the middle point between the two drains, it would remain at the same point if we opened both drains at the same time. If we put the ball at a different point or if one of the drains was bigger than the other, the ball would end up at the closer or bigger one. The region of the space that makes the ball tend to a certain attractor is called **basin of attraction** of the corresponding attractor. The boundaries of the basins of attraction have **saddle points** where the ball can "decide" where to go. At the boundaries, it is only at these saddle points where the ball can decide which region to go to (Fig. 2.7). We could also imagine a saddle point as the peak of a mountain; if the ball deviated a little, it would fall to one of the sides.

With this idea in mind, in order to model a particular differentiation process, one would need to find an expression for such a potential function. This potential function would need to have as many attractors as types of cells are involved in the process and its critical points will depend on parameters that are related to the biological signals that control the process. The potential is therefore characterised by the ways in which the attractors appear and disappear, i.e. the allowed bifurcations, and how the biological signals are maped into the mathematical parameters. Catastrophe theory provides the necessary tools to understand the singularities or bifurcations that can appear in a potential function.

In this section we will provide some basic concepts in this field. We will introduce definitions and important results for families of smooth functions and will state a very important result in catastrophe theory, *Thom's Classification Theorem*, which will allow us to introduce the elementary catastrophes, some of which will be

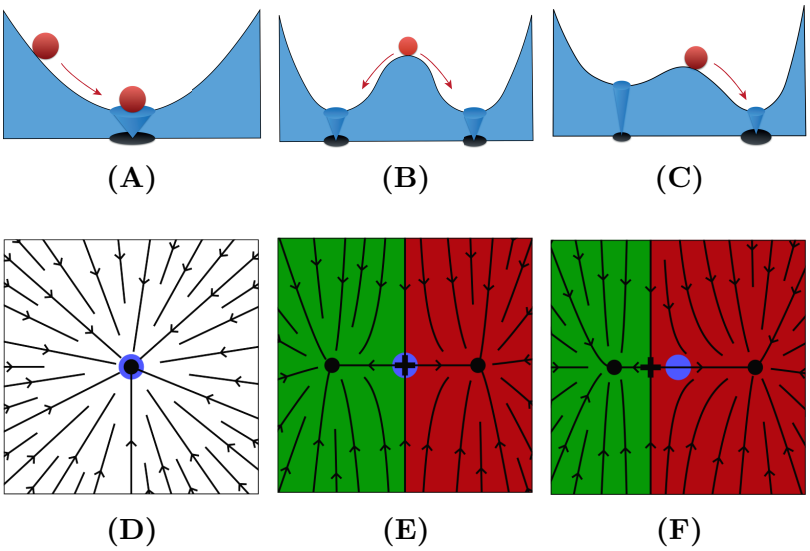


Figure 2.7: *Geometric view of the cell fate specification. (A-C) represent the simile of the sink, the red ball represents the state of the cell and the drains are the attractors. (D-F) represent the flow as if we looked at the sink from above. Adapted from (Corson & Siggia 2012). (A) The flow makes the ball to go towards the attractor. (B) At the saddle points, the cell can decide towards which attractor to go. (C) A signal can make the flow stronger towards a certain attractor. We can imagine that the drain gets bigger due to the signal, and the ball tends to that drain. (D-F) represent the same configurations as in (A-C) but as a flow field in the plane. The blue circle is the state of the cell, the black circles are the attractors and the saddles are shown as crosses. The basins of attraction for each fate are shown as green and red regions.*

used in Chapter 3. Most of the concepts summarised in this section are taken from the book Poston & Stewart (2014). We will not provide details about the proofs of the lemmas or theorems since we do not think they are strictly necessary for the understanding of the ideas presented in this thesis. We refer the reader to Poston & Stewart (2014) if interested in more details.

2.3.2 Families of functions and Thom's classification theorem

Let V be a family of functions

$$V : S \times C \to \mathbb{R}$$

where S is a manifold in $\mathbb{R}^n$ and C is another manifold in $\mathbb{R}^r$. Let us call $\mathbb{R}^n$ the **state space** and $\mathbb{R}^r$ the **control space**. For a fixed value c in C, we will denote by $V_c : S \to \mathbb{R}$ the potential function $V_c(x) = V(x, c)$ for $x \in S$.

For a fixed value c for the **control variables** or **parameters**, the function $V_c(x)$ will have certain critical points. Those will be given by the points $\boldsymbol{u}$ such that

$D_{\boldsymbol{x}}V(\boldsymbol{u}) = 0$. The critical point is **degenerate** if the rank of the Hessian matrix of V evaluated at $\boldsymbol{u}$ is not maximum, i.e. $det(D^2_{\boldsymbol{x}}V(\boldsymbol{u})) = 0$. If the critical point is not degenerate, we say that it is **non-degenerate**. We are interested in knowing how the critical points change depending on the value c, the parameters.

Let us define the **catastrophe manifold** $\mathcal{M}$ as the subset of $\mathbb{R}^n \times \mathbb{R}^r$ defined by:

$$\mathcal{M} = \{(x, c) \in \mathbb{R}^n \times \mathbb{R}^r : D_X V_c(x) = 0\} \tag{2.25}$$

where D_X is the gradient in the X variables. In other words, $\mathcal{M}$ is the set of points (x, c) such that x is a **critical point** of the function V_c.

The **catastrophe map** χ is the restriction to $\mathcal{M}$ of the natural projection

$$\begin{gathered} \pi : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^r \\ \pi(x, C) = C \end{gathered} \tag{2.26}$$

The **singularity set** $\mathcal{S}$ is the set of singular points in $\mathcal{M}$ at which χ is singular, that is, where the rank of the derivative $D\chi$ is less than r. Actually, it is not hard to show that $\mathcal{S}$ is the set of points $(x, c) \in \mathcal{M}$ at which $V_c(x)$ has a degenerate critical point. The image $\chi(\mathcal{S})$ in C is called the **bifurcation set**. It follows that $\mathcal{B}$ is the set on which the number and nature of the critical points change.

Summing up, the catastrophe manifold gives a description of the critical points for the family of functions $V_c(x)$, and thus it describes how they change as the parameters change. The bifurcation set describes the set of parameters for which important changes in the critical points take place.

Let us introduce a definition that allows to compare two families of functions.

Definition 1 *Two families of functions $V, W : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ are said to be **equivalent** around 0 if there exist:*

1. *a diffeomorphism $e : \mathbb{R}^r \to \mathbb{R}^r$;*
2. *a smooth map $y : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}^n$ such that for each $c \in \mathbb{R}^r$ the map $y_c : \mathbb{R}^n \to \mathbb{R}^n$ with $y_c(x) = y(x, c)$ is a diffeomorphism;*
3. *a smooth map $\gamma : \mathbb{R}^r \to \mathbb{R}$*

defined in a neighbourhood of 0, such that $G(x, c) = F(y_c(x), e(c)) + \gamma(c)$ for all $(x, c) \in \mathbb{R}^n \times \mathbb{R}^r$ in that neighbourhood.

Note that one can use this definition of equivalence around any point. The geometric meaning of this definition is that if V and W are equivalent around a point p then one can smoothly *deform* $\mathbb{R}^r$ and $\mathbb{R}^n$ so that F around p is the same as G. This equivalence is very important for understanding Thom's theorem.

The following lemmas are also very important in catastrophe theory, and allow to simplify the degeneracy of a critical point in a family of functions.

Theorem 1 (Splitting lemma for families) *Let $V : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ be smooth. Denote a point in $\mathbb{R}^n \times \mathbb{R}^r$ by $(x, c) = (x_1, \ldots, x_n, c_1, \ldots, c_r)$. Suppose that the hessian*

$$H = \left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_{1 \leq i,j \leq n}$$

has corank m at $(x, c) = 0$. Then V is equivalent to a family of the form

$$\tilde{V}(y_1(x, c), \ldots, y_m(x, c)) \pm y_{m+1}^2 \pm \cdots \pm y_n^2$$

Corollary 1.1 (Morse Lemma for families) *Let $V : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$ be smooth. Suppose that the Hessian*

$$\left(\frac{\partial^2 V}{\partial x_i \partial x_j} \right)_{1 \leq i,j \leq n}$$

has full rank at $(x, c) = 0$. Then V is equivalent (in the above sense) to a family of the form $\pm y_1^2 \pm \cdots \pm y_r^2$, said to be a Morse function[8]*.*

The Splitting Lemma for families of functions says that any family of functions V at critical point (x, c) is equivalent to a family of functions in which the degeneracy of the point is encoded in just m of the variables (the *essential* variables $y_1, \ldots, y_m$). It is also important to note that the parameter c no longer appears in the Morse part of the function. By Morse Lemma, if a family of functions $V_c(x)$ is Morse around a point, then it is of a fixed Morse type, independently of c.

With these definitions we can now introduce Thom's classification theorem.

2.3.2.1 Thom's Classification Theorem

Thom classifies all the different kinds of singularities that one can find in a family of functions with at most 5 parameters.

Theorem 2 (Thom's Classification Theorem) *Consider an r-parameter family of smooth functions $V : \mathbb{R}^n \times \mathbb{R}^r \to \mathbb{R}$, with $r \leq 5$ and suppose that V_{c_0} has a*

[8] Morse functions only have non-degenerate critical points

critical point at x_0. Then typically the restriction of the family to a neighbourhood of (x_0, c_0) in $\mathbb{R}^n \times \mathbb{R}^r$ is structurally stable and is equivalent to one of the following forms:

Non critical $W(u) = u_1$

Nondegenerate critical, or Morse $W(u) = u_1^2 + ... + u_i^2 - u_{i+1}^2 - ... - u_n^2$ *where* $0 \leq i \leq n$

Cuspoid catastrophes

1. *The fold (A_2): $W(u,t) = u_1^3 + t_1 u_1 + (M)$;*
2. *The cusp (A_3): $W(u,t) = \pm(u_1^4 + t_2 u_1^2 + t_1 u_1) + (M)$;*
3. *The swallowtail (A_4): $W(u,t) = u_1^5 + t_3 u_1^3 + t_2 u_1^2 + t_1 u_1 + (M)$;*
4. *The butterfly (A_5): $W(u,t) = \pm(u_1^6 + t_4 u_1^4 + t_3 u_1^3 + t_2 u_1^2 + t_1 u_1) + (M)$;*
5. *The wigwam (A_6): $W(u,t) = u_1^7 + t_5 u_1^5 + t_4 u_1^4 + t_3 u_1^3 + t_2 u_1^2 + t_1 u_1 + (M)$;*

Umbilic catastrophes

1. *The elliptic umbilic (D_4^-): $W(u,t) = u_1^2 u_2 - u_2^3 + t_3 u_1^2 + t_2 u_2 + t_1 u_1 + (N)$;*
2. *The hyperbolic umbilic (D_4^+): $W(u,t) = u_1^2 u_2 + u_2^3 + t_3 u_1^2 + t_2 u_2 + t_1 u_1 + (N)$;*
3. *The parabolic umbilic (D_5): $W(u,t) = \pm(u_1^2 u_2 + u_2^4 + t_4 u_2^2 + t_3 u_1^2 + t_2 u_2 + t_1 u_1) + (N)$;*
4. *The second elliptic umbilic (D_6^-): $W(u,t) = u_1^2 u_2 - u_1^5 + t_5 u_2^3 + t_4 u_2^2 u_+ t_3 u_1^2 + t_2 u_2 + t_1 u_1 + (N)$;*
5. *The second hyperbolic umbilic (D_6^+): $W(u,t) = u_1^2 u_2 + u_1^5 + t_5 u_2^3 + t_4 u_2^2 u_+ t_3 u_1^2 + t_2 u_2 + t_1 u_1 + (N)$;*
6. *The symbolic umbilic (E_6): $W(u,t) = \pm(u_1^3 + u_2^4 + t_5 u_1 u_2^2 + t_4 u_2^2 + t_3 u_1 u_2 + t_2 u_2 + t_1 u_1) + (N)$;*

Here $(u_1, \ldots, u_n) \in \mathbb{R}^n$, $(t_1, \ldots, t_r) \in \mathbb{R}^r$; the symbol (M) indicates a Morse function of the form $u_2^2 + \cdots + u_i^2 - u_{i+1}^2 - \cdots - u_n^2$ $(2 \leq i \leq n)$ and similarly (N) indicates the function $u_3^2 + \cdots + u_i^2 - u_{i+1}^2 - \cdots - u_n^2$ $(3 \leq i \leq n)$.

In what follows of this section we are going to study some important catastrophes from Thom's Classification Theorem. In particular, we will focus on the fold and the cusp since we will use them in Chapter 3.

2.3.3 Study of Fold and Cusp Catastrophes

2.3.3.1 Fold Catastrophe

The fold is the most basic catastrophe and it has the form:

$$V(x,c) = V_c(x) = x^3 + cx. \tag{2.27}$$

It is a cubic equation with graph always crossing the point $x = 0$. As we will see, depending on the value of the control parameter c, the potential function V_c will have two, one or no critical points (Figure 2.8).

The catastrophe manifold $\mathcal{M}$ is given by the critical points of the potential function for each value of c:

$$0 = \frac{dV_c(x)}{dx} = 3x^2 + c. \tag{2.28}$$

Therefore we can use the x-coordinate as a chart for $\mathcal{M}$ and write it as

$$\mathcal{M} = \{(x,c) \in \mathbb{R} \times \mathbb{R} : c = -3x^2\} = \{(x, -3x^2) \in \mathbb{R} \times \mathbb{R}_{\leq 0}\}. \tag{2.29}$$

This means that for each value of c, the potential function has two critical points ($c < 0$), one critical point ($c = 0$) or no critical points ($c > 0$) (See Figure 2.9).

If we look at the stability of such critical points,

$$D^2V(x, -3x^2) = \frac{d^2V}{dx^2}(x, -3x^2) = 6x, \tag{2.30}$$

which sign depends on the sign of x. If $x < 0$, the critical point is stable. If $x > 0$, the critical point is unstable. And if $x = 0$ the critical point is degenerate. In fact, the singularity set is the set of points $(x,c) \in \mathcal{M}$ such that

$$0 = D^2V(x,c) = D^2V(x, -3x^2) = \frac{d^2V}{dx^2}(x, -3x^2) = 6x \iff x = c = 0. \tag{2.31}$$

Hence, the singularity set $\mathcal{S}$ of the fold catastrophe is the point $(0,0)$ and the bifurcation set $\mathcal{B}$ is the point $c = 0$ in the control space.

The catastrophe manifold helps us to identify regions in the control space for which the potential function will have the same critical points. For example, for the fold catastrophe, if $c < 0$, the potential will contain two critical points ($x_1^* = -\sqrt{-c}/3$ and $x_2^* = \sqrt{-c}/3$); if $c = 0$, the potential has one degenerate critical point ($x^* = 0$); and if $c > 0$ the potential function $V_c(x)$ contains no critical points (see Figure 2.8).

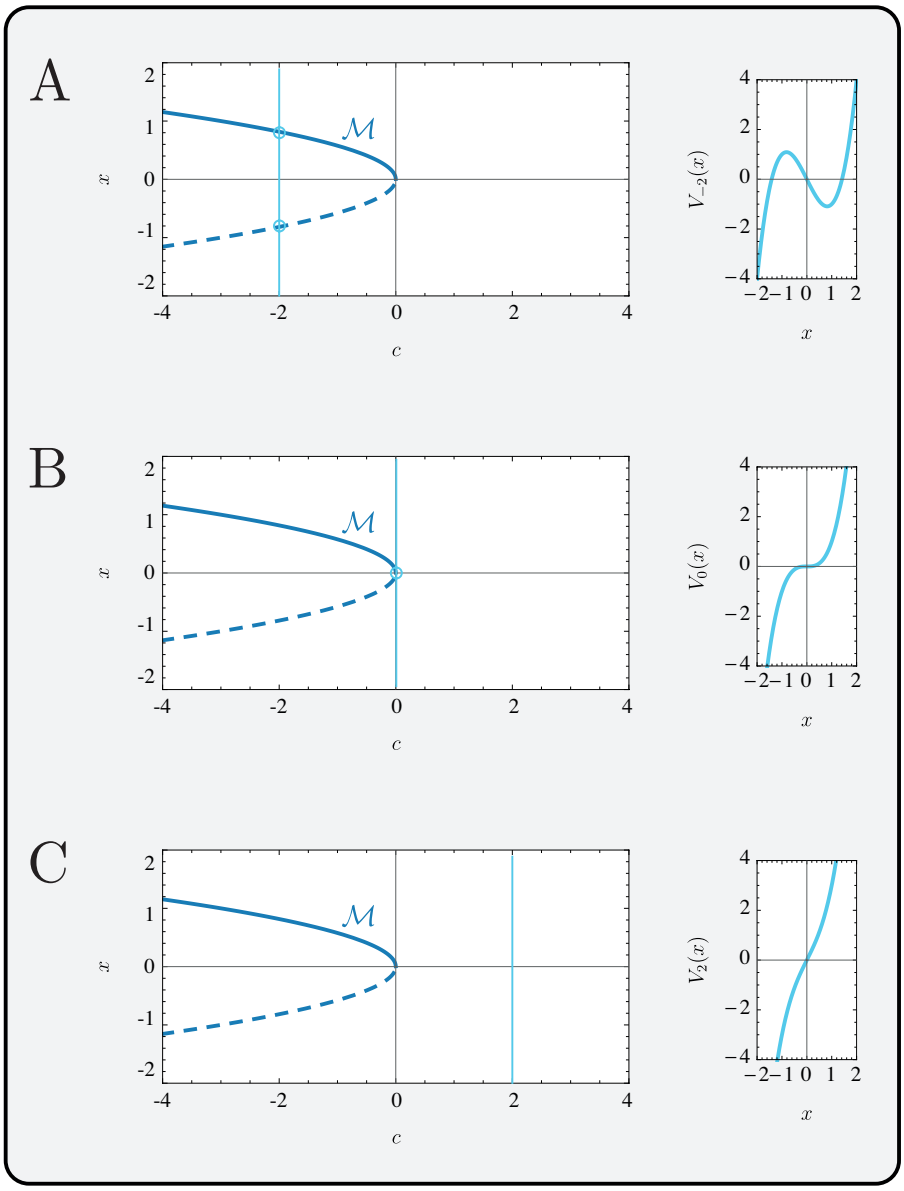


Figure 2.8: *On the left, bifurcation diagram or catastrophe manifold of the fold catastrophe. On the right, plot with the fold potential for the corresponding value of the control parameter c. Continuous line in the bifurcation diagram represents stable points, dashed line represents unstable points. (A) In this case $c = -2$, for which the potential has two critical points, a minimum and a maximum. (B) In this case $c = 0$ and corresponds to a bifurcation point. The potential function only contains a degenerate critical point at $x = 0$. (C) In this case $c = 2$ and the potential function does not contain any critical point.*

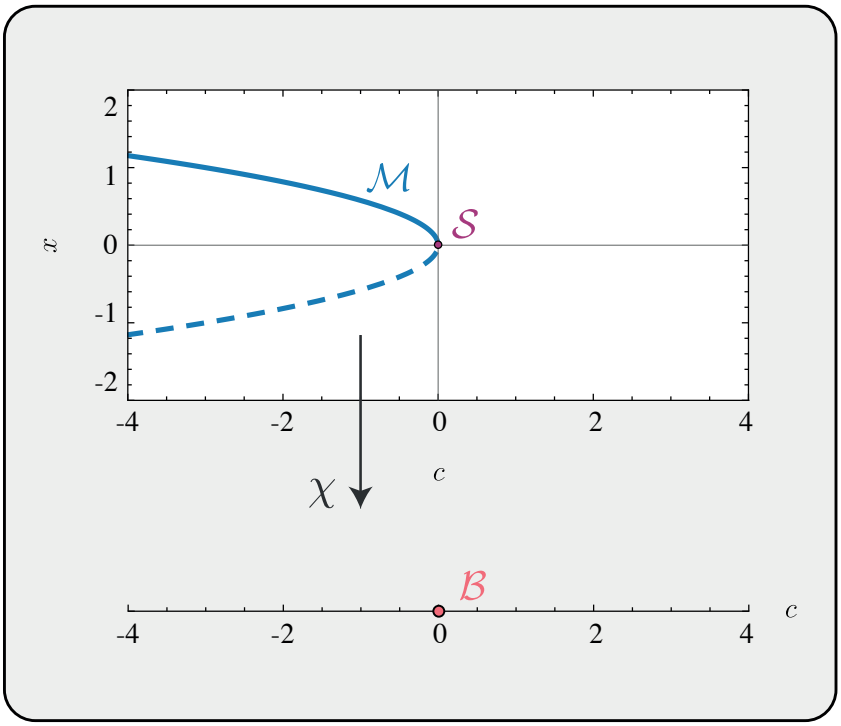


Figure 2.9: *Catastrophe manifold $\mathcal{M}$, catastrophe map χ, singularity set $\mathcal{S}$ and bifurcation set $\mathcal{B}$ of the fold catastrophe given in Equation 2.29. Continuous line represents stable points, dashed line represents unstable points.*

2.3.3.2 Cusp Catastrophe

The cusp catastrophe is one of the most famous ones and Zeeman used of in many applications, see Zeeman (1976) (some of them becoming very controversial, however (Guckenheimer 1978)). This catastrophe has the form

$$V_{a,b}(x) = x^4 + ax^2 + bx, \tag{2.32}$$

which is a polynomial of order 4 in one variable. In this case, the control space is 2-dimensional on a, b, and depending on their values, this potential function will contain one, two or three critical points (see Figure 2.10).

The catastrophe manifold $\mathcal{M}$ is given by the points in $\mathbb{R} \times \mathbb{R}^2$ such that

$$0 = \frac{dV_{a,b}(x)}{dx} = 4x^3 + 2ax + b = 0. \tag{2.33}$$

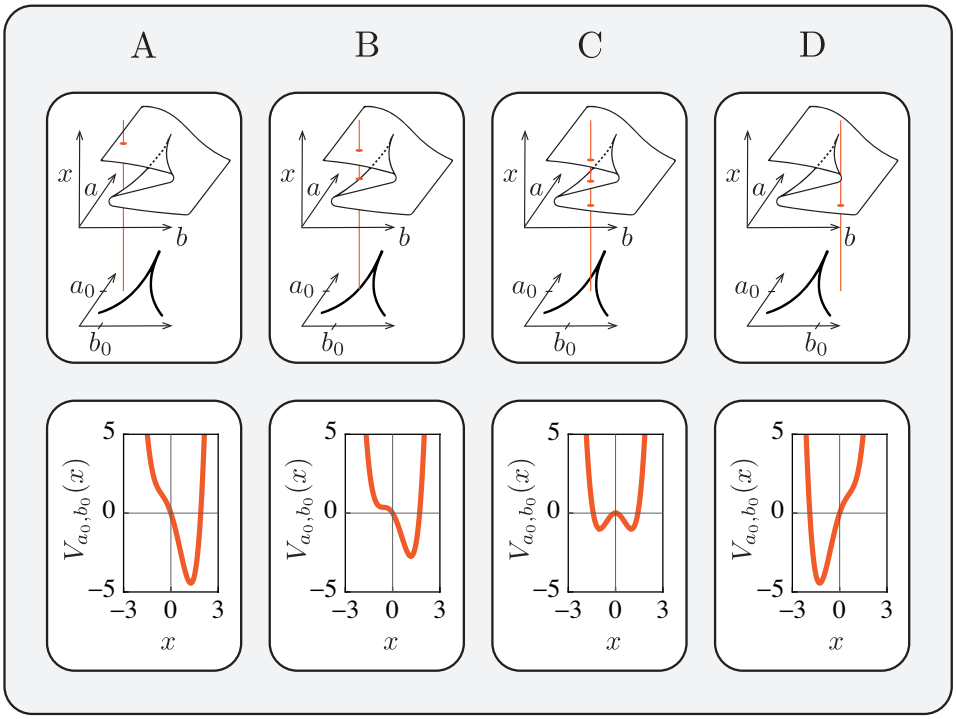


Figure 2.10: *Four different potential functions from the cusp family for control parameters in different regions of the control space. Each row of plots (A-D) shows, on the top figure, the catastrophe manifold intersected with the line (in orange)* $L_{a_0,b_0} = \{(x, a_0, b_0) \in \mathbb{R} \times \mathbb{R}^2 : x \in \mathbb{R}\}$ *for a fixed value* $a_0 < 0$ *and different values of* b_0*, displaying the critical points (orange ellipses); and, on the bottom figure, the potential function corresponding to such values of* a *and* b *with the predicted critical points. (A) The control parameters are such that* $8a_0^3 + 27b_0^2 > 0$ *and* $b_0 < 0$*, therefore the potential has only one critical point with* $x > 0$*. (B) In this case* $(a_0, b_0) \in \mathcal{B}$ *and* $b_0 > 0$*, so the potential has two critical points, one of them is degenerate. (C) The control parameters are such that* $8a_0^3 + 27b_0^2 < 0$ *so the potential function has three critical points. (D) Similar to (A), The control parameters are such that* $8a_0^3 + 27b_0^2 > 0$ *but now* $b_0 > 0$*, therefore the potential has one, now negative, critical point.*

Hence, we can use (x, a) as a chart on $\mathcal{M}$ and write b in terms of x and a as $b = -4x^3 - 2ax$. This allows us to write the catastrophe manifold as the set of points

$$\mathcal{M} = \{(x, a, b) \in \mathbb{R} \times \mathbb{R}^2 : b = -4x^3 - 2ax\} = \{(x, a, -4x^3 - 2ax) \in \mathbb{R} \times \mathbb{R}^2\}. \quad (2.34)$$

Figure 2.11 contains a plot of this two dimensional manifold in $\mathbb{R}^3$.

The singularity set $\mathcal{S}$ of the cusp catastrophe is given by the points in the catastrophe manifold such that

$$D^2V(x, a, -4x^3 - 2ax) = \frac{d^2V}{dx^2}(x, a, -4x^3 - 2ax) = 12x^2 + 2a = 0. \quad (2.35)$$

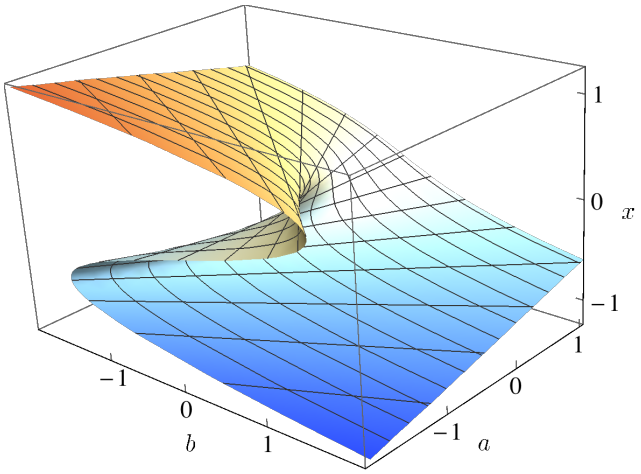


Figure 2.11: *Catastrophe manifold $\mathcal{M}$ of the cusp catastrophe.*

Therefore, $a = -6x^2$; and since $b = -4x^3 - 2ax$ in $\mathcal{M}$, then

$$\mathcal{S} = \{(x, a, b) \in \mathcal{M} : a = -6x^2\} = \{(x, -6x^2, 8x^3) \in \mathbb{R} \times \mathbb{R}^2, x \in \mathbb{R}\}, \tag{2.36}$$

which is a curve in $\mathbb{R}^3$ (see Figure 2.12).

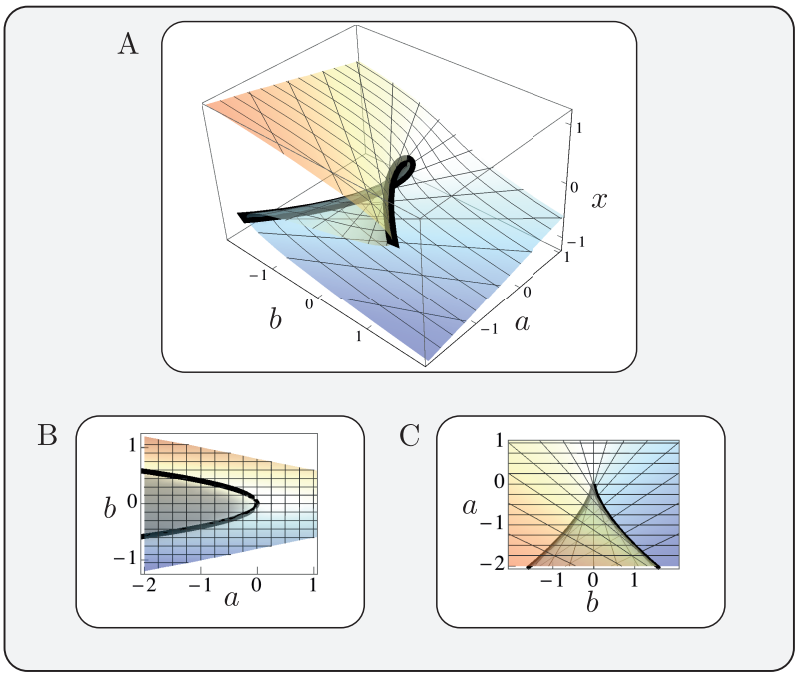


Figure 2.12: *Different viewpoints of the catastrophe manifold $\mathcal{M}$ and singularity set $\mathcal{S}$ (black thick line) of the cusp catastrophe.*

Finally, the bifurcation set $\mathcal{B}$ is given by the image of this set by the catastrophe map, which is the projection of this curve on the control space:

$$\mathcal{B} = \{(-6t^2, 8t^3) \in \mathbb{R}^2, t \in \mathbb{R}\} = \{(a, b) \in \mathbb{R}^2 : \Delta(a, b) = 8a^3 + 27b^2 = 0\}. \quad (2.37)$$

As we see in Figure 2.13, the bifurcation set defines three regions in the control space. For values of the control parameters (a_0, b_0) in the region such that $\Delta(a_0, b_0) = 8a_0^3 + 27b^2 < 0$, the potential function contains three critical points (two stable and one unstable). If (a_0, b_0) is in the bifurcation set, the potential function contains two critical points (one stable and one degenerate). And if the control parameters are in the region such that $\Delta(a_0, b_0) > 0$, then the potential function contains just one stable critical point (see Figure 2.10). The coordinates of the critical points in each case can be computed with Cardano's formula (see Appendix A).

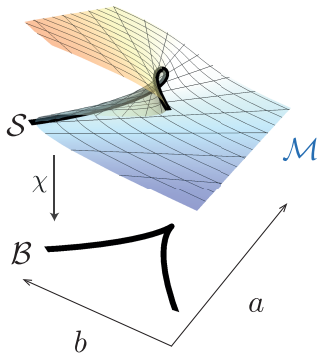


Figure 2.13: *Catastrophe manifold $\mathcal{M}$, catastrophe map χ, singularity set $\mathcal{S}$ and bifurcation set $\mathcal{B}$ of the cusp catastrophe given in Equations 2.34.*

We now move to methods of parameter estimation, that we use in Chapter 4.

2.4 Parameter estimation

2.4.1 Bayesian methods

Suppose that we have a continous random variable $\boldsymbol{X}(\boldsymbol{\theta})$ that models a certain process, and the outcome of which depends on the value of the parameter vector $\boldsymbol{\theta}$. Normally, one would have some data $\boldsymbol{X}_0$ about the process of interest and would like to find the parameter(s) with which the model can reproduce such data. Bayesian methods tackle this problem by treating the parameters as random variables and looking at the probability that a certain parameter vector $\boldsymbol{\theta}$ reproduces data $\boldsymbol{X}_0$. They build upon Bayes' theorem, which states

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0) = \frac{\pi(\boldsymbol{X}_0 \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int \pi(\boldsymbol{X}_0 \mid \boldsymbol{\theta}')\pi(\boldsymbol{\theta}')d\boldsymbol{\theta}'}. \tag{2.38}$$

The function $\pi(\boldsymbol{X}_0 \mid \boldsymbol{\theta})$ is the **likelihood function** and represents the probability of obtaining data $\boldsymbol{X}_0$ from our model when the parameters take value $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$. The function $\pi(\boldsymbol{\theta})$ is the **prior distribution** of the parameters, and it contains our prior *knowledge* or *belief* regarding the parameter vector $\boldsymbol{\theta}$ before seeing the data. The denominator can be regarded as the probability of the data, and it is essentially a proportionality constant. The quantity of interest is the **posterior distribution** $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0)$, which contains the updated information about the parameters taking the data into consideration. We can rewrite Equation 2.38 as

$$\begin{array}{ccccc} \pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0) & \propto & \pi(\boldsymbol{\theta}) & \times & \mathcal{L}(\boldsymbol{X}_0 \mid \boldsymbol{\theta}), \\ \text{posterior} & & \text{prior} & & \text{likelihood} \end{array} \tag{2.39}$$

i.e. the posterior combines the prior knowledge about the parameters and the observed data to give the true distribution of the parameters.

When an analytical expression of the posterior distribution can be obtained, one can compute important quantities related to the parameters, such as its moments, marginal and conditional distributions. For example, the mean of the posterior distribution is given by

$$\mathbb{E}[\boldsymbol{\theta}] = \int \boldsymbol{\theta}\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0)d\boldsymbol{\theta}.$$

It is often the case that one is unable to obtain posterior distributions in closed form, particularly when more realistic processes are modelled. To solve this problem, Markov chain Monte Carlo (MCMC) methods such as the Gibbs sam-

pler and the Metropolis-Hastings algorithm or Aproximate Bayesian computation (ABC) methods give a way of approximating such posterior distribution, and in turn study properties of the parameters that reproduce the data. MCMC methods rely on the likelihood function being tractable, while ABC methods are developed to approximate posterior distributions in the case of intractable likelihoods.

In the following subsections we will give an overview of such methods; in particular, we will focus on the adaptive Metropolis (Haario et al. 2001) and ABC SMC (Toni et al. 2009) algorithms since they will be important for the methodology developed in Chapter 4.

2.4.2 Markov chain Monte Carlo (MCMC) methods

Markov chain Monte Carlo (MCMC) methods provide a way of generating samples from a distribution π, the target distribution, which is known up to a normalisation constant. In our case, π will be the posterior distribution of the parameters $\boldsymbol{\theta}$, given the model $\boldsymbol{X}(\boldsymbol{\theta})$ and the data $\boldsymbol{X}_0$. For the simulation of the distribution π, MCMC methods generate an ergodic Markov chain $(\boldsymbol{\theta}^{(t)})$ with **stationary distribution** π. This means that, starting from an arbitrary value of the parameter vector $\boldsymbol{\theta}^{(0)}$, MCMC algorithms generate a chain $(\boldsymbol{\theta}^{(t)})$ that converges in distribution to the target distribution π (i.e. there will be a T such that $\boldsymbol{\theta}^{(t)} \sim \pi$ for all $t > T$). Therefore, one can approximate the distribution π by using the sampled values in the chain from $t > T$.

There is a wide range of MCMC algorithms, but we will focus on the Metropolis-Hastings (MH) algorithm (Hastings 1970, Metropolis et al. 1953) and a variant of it called adaptive Metropolis (AM) algorithm (Haario et al. 2001), since the later one will be mentioned in Chapter 4. For more details on MCMC methods we refer the reader to Christian & Casella (1999) and Brooks et al. (2011).

2.4.2.1 Metropolis-Hastings algorithm

The Metropolis-Hastings (MH) algorithm was first introduced by Hastings (1970). The method is given in Algorithm 2. As other MCMC methods, the algorithm starts from an initial parameter vector $\boldsymbol{\theta}^{(0)}$ and creates a chain $\{\boldsymbol{\theta}^{(t)}\}$ that converges to the distribution of interest π. The chain is created by generating new values of the parameter $\boldsymbol{\theta}^*$ using a **proposal distribution** $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right)$. The parameter value $\boldsymbol{\theta}^*$ is accepted or rejected based on the so called **acceptance probability** $\alpha\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}\right)$, which is a function of the prior and likelihood distributions evaluated at the corresponding parameter values.

Algorithm 2: The Metropolis-Hastings algorithm

Initialise $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_m^{(0)})$
for *t = 1, 2, ..., N* **do**
  Given $\boldsymbol{\theta}^{(t-1)}$, draw a candidate value $\boldsymbol{\theta}^*$ from the proposal density $q(\cdot \mid \boldsymbol{\theta}^{(t-1)})$
  Calculate $\alpha\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}\right) = \min\left\{1, \frac{\pi\left(\boldsymbol{\theta}^* \mid \boldsymbol{X}_0\right) q\left(\boldsymbol{\theta}^{(t-1)} \mid \boldsymbol{\theta}^*\right)}{\pi\left(\boldsymbol{\theta}^{(t-1)} \mid \boldsymbol{X}_0\right) q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right)}\right\}$
  Draw $u \sim \mathcal{U}([0,1])$
  if $u \leq \alpha\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}\right)$ **then**
    Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$
  else
    Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$
  end
end

The MH algorithm relies on the evaluation of the likelihood function and the choice of the proposal distribution, which is critical for its convergence.

The simplest MH algorithm is such that the proposal distribution only depends on the first argument, i.e. $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right) = q\left(\boldsymbol{\theta}^*\right)$, which is known as the **independent Metropolis-Hastings** algorithm. Another common choice is $q\left(\boldsymbol{\theta}^* \mid \boldsymbol{\theta}^{(t-1)}\right) = q\left(\left|\boldsymbol{\theta}^* - \boldsymbol{\theta}^{(t-1)}\right|\right)$, which defines the *random walk Metropolis-Hastings* and it was proposed by Metropolis in 1953. The choice of this proposal distribution, as previously said, is important since it might lead to a very low or high acceptance probability affecting the convergence of the chain.

For this reasons Haario et al. (2001), among others, developed a variant of the MH algorithm called **adaptive Metropolis** (AM), that automatically updates the proposal distribution taking into account the information about the chain accumulated so far.

2.4.2.2 An adaptive Metropolis algorithm

The adaptive Metropolis (AM) algorithm introduced in Haario et al. (2001) adapts its proposal distribution along the way according to the covariance matrix of the past steps of the chain. The method follows Algorithm 3.

The method can be divided into two parts. For $t = 1, ..., N_1$, the algorithm works as a normal MH algorithm with Gaussian proposal distribution with mean

Algorithm 3: An adaptive Metropolis algorithm

Initialise $\boldsymbol{\theta}^{(0)} = (\theta_1^{(0)}, \ldots, \theta_m^{(0)})$
for $t = 1, 2, \ldots, N$ **do**
  if $t < N_1$ **then**
    Given $\boldsymbol{\theta}^{(t-1)}$, draw a candidate value $\boldsymbol{\theta}^*$ from the proposal density $\mathcal{N}(\boldsymbol{\theta}^{(t-1)}, C_0)$
  else
    Compute $C_t = s_d \text{cov}(\boldsymbol{\theta}^{(0)}, \boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(t-1)}) + s_d \varepsilon I_d$
    Given $\boldsymbol{\theta}^{(t-1)}$, draw a candidate value $\boldsymbol{\theta}^*$ from the proposal density $\mathcal{N}(\boldsymbol{\theta}^{(t-1)}, C_t)$
  end
  Calculate $\alpha\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}\right) = \min\left\{1, \frac{\pi\left(\boldsymbol{\theta}^* | \boldsymbol{X}_0\right)}{\pi\left(\boldsymbol{\theta}^{(t-1)} | \boldsymbol{X}_0\right)}\right\}$
  Draw $u \sim \mathcal{U}\left([0,1]\right)$
  if $u \leq \alpha\left(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{(t-1)}\right)$ **then**
    Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^*$
  else
    Set $\boldsymbol{\theta}^{(t)} = \boldsymbol{\theta}^{(t-1)}$
  end
end

$\boldsymbol{\theta}^{(t-1)}$ and fixed to an arbitrary, strictly positive definite initial covariance matrix C_0 (which is selected according to one's best prior knowledge). For $t > N_1$, the covariance matrix C_t for the Gaussian proposal distribution is computed from the previous steps of the chain (see Algorithm 3), therefore, taking into account the information accumulated in the chain, making it more efficient. However, it is important to note that the chain is no longer Markovian since the acceptance probability of $\boldsymbol{\theta}^{(t)}$, for $t > N_1$, does not only depend on $\boldsymbol{\theta}^{(t-1)}$ but on the whole past of the chain. This implies that additional conditions need to be satisfied in order to ensure that the chain converges to the target distribution (Brooks et al. 2011).

2.4.3 Approximate Bayesian computation scheme: ABC methods

Approximate Bayesian Computation (ABC) methods, also known as likelihood-free methods, have been developed for inferring the posterior distribution of the parameters when the likelihood function is either too complex or too expensive to compute

but observations of the model can be simulated fairly easily. ABC methods have appeared as a very satisfactory approach to intractable likelihood problems for the past eighteen years (Marin et al. 2012), being applied to a diverse range of problems, including population genetics (Beaumont et al. 2002), pathogen transmission (Tanaka et al. 2006) and reaction network models (Toni & Stumpf 2010, Liepe et al. 2014) among others.

These methods use a comparison between the experimental and simulated data to measure the goodness of fit, instead of using the likelihood function.

The term *approximate Bayesian computation* was first used in Rubin et al. (1984), where he proposed Algorithm 4. For a given observation $\boldsymbol{X}_0$, the algorithm generates parameter values $\boldsymbol{\theta}^*$ from the prior distribution and $\boldsymbol{X}(\Theta^*)$ from the model until $\boldsymbol{X}(\Theta^*)$ equals the observation $\boldsymbol{X}_0$.

Algorithm 4: Likelihood-free rejection sampler 1

```
for i = 1, 2, ..., N do
    repeat
        Generate θ* from the prior distribution π(·)
        Generate X(θ') from the model
    until X(θ*) = X_0;
    Set θ_i = θ*
end
```

It is clear that the parameters $(\boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \ldots, \boldsymbol{\theta}_N)$, obtained with Algorithm 4, are an independent and identically distributed (iid) sample from the desired posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0)$. However, this algorithm has a very high rejection rate because it requires exact match between the simulated and observed data. In consequence, it is not applicable to real problems.

Pritchard et al. (1999) generalised this method, who proposed the first authentic ABC algorithm, sumarised in Algorithm 5. For a given observation $\boldsymbol{X}_0$, the algorithm generates parameter values $\boldsymbol{\theta}^*$ from the prior distribution and $\boldsymbol{X}(\Theta^*)$ from the model. It then compares a summary statistic of the observation $S(\boldsymbol{X}_0)$ and the simulated data $S(\boldsymbol{X}(\Theta^*))$, and accepts the sampled parameter value if these summary statistics are sufficiently *similar*. This similarity is measured by a distance function d and a tolerance level ε.

In contrast to Rubin's algorithm, Algorithm 5 does not sample from the posterior but from an *approximation* of the posterior that, if the summary statistic S is representative enough and the tolerance level ε is small enough, should be a

good approximation of the posterior distribution (Marin et al. 2012).

Algorithm 5: Likelihood-free rejection sampler 2

```
for i = 1, 2, ..., N do
    repeat
        Generate θ* from the prior distribution π(·)
        Generate X(θ*) from the model
    until d(S(X(θ*)), S(X_0)) ≤ ε;
    Set θ_i = θ*
end
```

Many different ABC algorithms such as ABC rejection sampler, ABC MCMC and sequential Monte Carlo ABC have been proposed since. They are all based on the idea of sampling parameter vectors $\boldsymbol{\theta}^*$ and retaining the ones which simulated data $\boldsymbol{X}(\boldsymbol{\theta}^*)$ are somewhat similar to the observed data $\boldsymbol{X}_0$. We refer the reader to Marin et al. (2012), Beaumont et al. (2009) for reviews on ABC methods and to Brooks et al. (2011) for more technical details. We will now focus on a particular ABC algorithm called sequential Monte Carlo ABC (ABC SMC), which will be used in Chapter 4.

2.4.3.1 Sequential Monte Carlo ABC (ABC SMC)

This version of the ABC algorithm *borrows* ideas from importance sampling and sequential Monte Carlo, as the name suggests. The ABC SMC was first proposed by (Sisson et al. 2007), and later corrected by (Beaumont et al. 2009, Sisson et al. 2009, Toni et al. 2009). Various ABC SMC algorithms are proposed in the literature (see for example Beaumont et al. (2009), Moral et al. (2012), Sisson et al. (2007), Toni et al. (2009)), but here we focus on the fairly general ABC SMC algorithm by Toni et al. (2009). The ABC SMC sampler methodology approximates a sequence of probability distributions $\{\pi_t\}_{0\leq t\leq T}$ that satisfies the condition that $d(\boldsymbol{X}(\boldsymbol{\theta})), \boldsymbol{X}_0) \leq \varepsilon_t$,

$$\{\pi_t\}_{0\leq t\leq T} = \{\pi(\boldsymbol{\theta} \mid d(\boldsymbol{X}(\boldsymbol{\theta})), \boldsymbol{X}_0) \leq \varepsilon_t)\}_{0\leq t\leq T}. \tag{2.40}$$

In order to produce these probability distributions, the algorithm starts by sampling parameter values $\boldsymbol{\theta}^*$ from a prior distribution $\pi(\boldsymbol{\theta})$. Then, the algorithm accepts N parameter values that satisfy $d(\boldsymbol{X}(\boldsymbol{\theta}^*)), \boldsymbol{X}_0) \leq \varepsilon_1$. It assigns a set of equal weights, $\{\omega_1^{(i)} = 1/N\}_{i=1}^N$, to each accepted parameter value $\boldsymbol{\theta}_1^{(i)}$, $i = 1, \ldots, N$, in

this context called particles.

Then the algorithm proceeds in a sequential manner. In each step t, a set of N particles $\{\boldsymbol{\theta}_t^{(i)}\}_{i=1}^N$ is generated. This is done by first sampling a particle $\boldsymbol{\theta}^{**}$ from the discrete distribution with support the finite set of particles generated in the previous step, $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$, and probabilities given by the weights $\{\omega_{t-1}^{(i)}\}_{i=1}^N$. Then, the sample is perturbed using the Markov kernel function $K_t(\boldsymbol{\theta}^*, \boldsymbol{\theta}^{**})$. These steps are repeated until N particles $\boldsymbol{\theta}^*$ that satisfy the conditions $\pi(\boldsymbol{\theta}^*) > 0$ and $d(\boldsymbol{X}(\boldsymbol{\theta}^*), \boldsymbol{X}_0) \leq \varepsilon_t$ are found.

The ABC SMC algorithm developed in (Toni et al. 2009) is provided in Algorithm 6.

Algorithm 6: ABC SMC algorithm

Initialize $\varepsilon_1, \ldots, \varepsilon_T$
for *i = 1, 2, ..., N* **do**
  repeat
    Generate $\boldsymbol{\theta}^*$ from the prior distribution $\pi(\cdot)$
  until $d(\boldsymbol{X}(\boldsymbol{\theta}^*), \boldsymbol{X}_0) \leq \varepsilon_t$;
  Set $\boldsymbol{\theta}_1^{(i)} = \boldsymbol{\theta}^*$ and set the weight $\omega_1^{(i)}$ for particle $\boldsymbol{\theta}_1^{(i)}$ equal to $1/N$.
end
for *t = 2, ..., T* **do**
  for *i = 1, 2, ..., N* **do**
    repeat
      Generate $\boldsymbol{\theta}^{**}$ from the previous population $\{\boldsymbol{\theta}_{t-1}^{(i)}\}_{i=1}^N$
      Perturb the particle to obtain $\boldsymbol{\theta}^* \sim K_t(\cdot, \boldsymbol{\theta}^{**})$, where K_t is a perturbation kernel.
    until $\pi(\boldsymbol{\theta}^*) > 0$ *and* $d(\boldsymbol{X}(\boldsymbol{\theta}^*), \boldsymbol{X}_0) \leq \varepsilon_t$;
    Set $\boldsymbol{\theta}_t^{(i)} = \boldsymbol{\theta}^*$ and calculate the weight for particle $\boldsymbol{\theta}_t^{(i)}$,
    $$\omega_t^{(i)} = \frac{\pi(\boldsymbol{\theta}_t^i)}{\sum_{j=1}^N \omega_{t-1}^{(j)} K_t\left(\boldsymbol{\theta}_{t-1}^{(j)}, \boldsymbol{\theta}_t^{(i)}\right)}$$
  end
  Normalise the weights.
end

The behaviour of the algorithm strongly depends on the decreasing sequence of $\{\varepsilon_t\}_{t=1}^T$ and the perturbation kernels $\{K_t(\cdot,\cdot)\}_{t=1}^T$ since they affect the acceptance ratio and thus the number of simulations that need to be computed (Filippi et al.

2013, Silk et al. 2012). However, (Filippi et al. 2013) shows that there are perturbation kernels which are more efficient than others, and (Silk et al. 2012) advises that it is better to reduce the tolerance levels in a rapid way otherwise the method can get stuck in wrong parameter regions. This will be discussed in Chapter 4.

Chapter 3

A new mathematical model for vulval development in *C. elegans*

3.1 Biological background

Caenorhabditis elegans (*C. elegans* in short) is a small (around 1.2 mm long and 70 μm in circumference) free-living soil nematode (roundworm) which feeds on microbes, primarily bacteria. As an adult, it consists of only about 1,000 somatic cells and 1,000-2,000 germ cells depending on the sex of the animal (which can be male or hermaphrodite).

Despite it is a relatively simple organism lacking many organs that other more complex organisms have, the worm is a good model organism that allows us to understand basic general questions of animal development (Alberts et al. 2002). Among other characteristics, its suitability for genetic analysis, the invariant lineage of its small number of cells, the transparency of the embryo and its short embryonic development, with life cycle of about 3 days under optimal conditions, stand out. This is why many biological fields such as genomics, cell biology, neuroscience and ageing focus their research on it.

Vulval development is a broadly studied stage in *C. elegans* development. The vulva is an adult structure that develops during the larval stages of the hermaphrodite worm. The mature vulva contains 22 cells of different types, and more than 40 genes are involved in its development. It is derived from six ectodermal cells (P3.p, P4.p, P5.p, P6.p, P7.p and P8.p). They are **vulval precursor cells** or VPCs; they are partially differentiated and are situated in a row along the antero-posterior axis in the ventral side of the larva. During vulval development, the VPCs can take three different fates: primary (1°), secondary (2°) and tertiary (3°). In other words, they can become those three different types of cells. The first two fates determine vulval cells, while the tertiary fate is non-vulval and cells taking this fate fuse with the

epidermis. In the wild type (WT) (i.e. the typical form of *C. elegans* as it occurs in nature), P6.p assumes primary fate, P5.p and P7.p assume secondary fate and the rest of VPC take tertiary fate (see Figure 3.1). This means that only three of them (P5.p, P6.p and P7.p) lead to the vulva, and the rest join the epidermis (Sternberg 2005, Lints & Hall 2009, Wolpert et al. 2015). P3.p often does not divide and fuses with the epidermis (Sternberg & Horvitz 1986, Sternberg 2005, Corson & Siggia 2012) assuming tertiary fate, therefore we will not include it in our study.

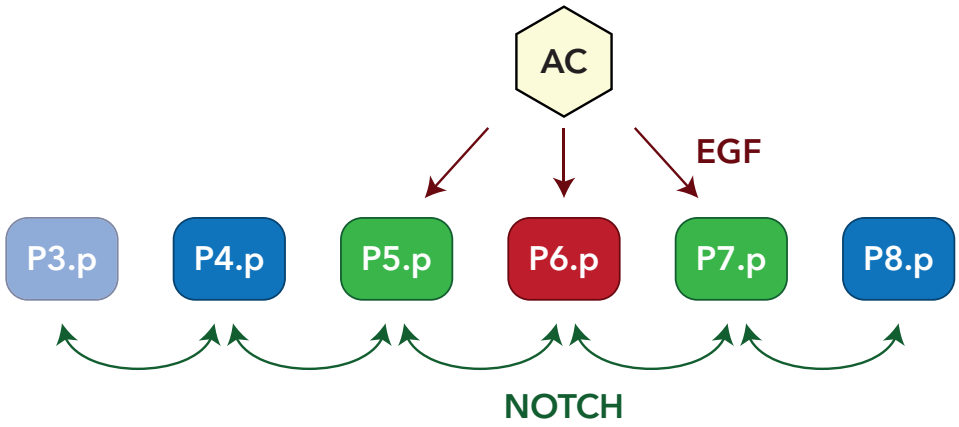


Figure 3.1: *Schematic representation of Vulval development in C. elegans, wild type (WT) pattern. Red colour represents the primary fate (P6.p), green colour represents secondary fate (P5.p and P7.p) and blue and shaded blue colours represents tertiary fate (P3.p, P4.p and P8.p). Signals are represented by arrows, being EGF signal represented by red arrows and Notch signal by green arrows. P3.p is coloured with a shaded blue because we will not include it in our study.*

These vulval precursor cells, however, are apparently equivalent in their ability to develop as vulval cells from the beginning, and how they determine their fates mainly relies upon the activation of two different signalling pathways (Schmid & Hajnal 2015) that are involved in the patterning mechanism: the EGF-Ras-MAP kinase pathway and the Delta-Notch pathway. The activation of the first signalling pathway depends on the EGF ligand, which is a signalling molecule secreted by the anchor cell. Whereas the delta ligands that activate the second pathway are secreted by the VPCs themselves. VPC response to external signals begins in the L2 larval stage and stops shortly after the first round of division in L3. The activation of the first pathway promotes the expression of the egl-17 gene, which is characteristic of primary fate cells; while the activation of the second pathway activates the expression of the lip-1 gene which is highly expressed in secondary fate VPCs (Sternberg & Horvitz 1989, Hoyos et al. 2011). Moreover, the two signalling pathways also interact through three intra and intercellular crosstalks. The phosphorylation of the

MAP kinase activates the synthesis of Delta ligands and downregulates the Notch receptor, while the activation of the Delta-Notch pathway inhibits MAPKP activity (see Figure 3.2). This crosstalk could be the reason why the pattern 3°2°1°2°3° is so precise (Fisher et al. 2007).

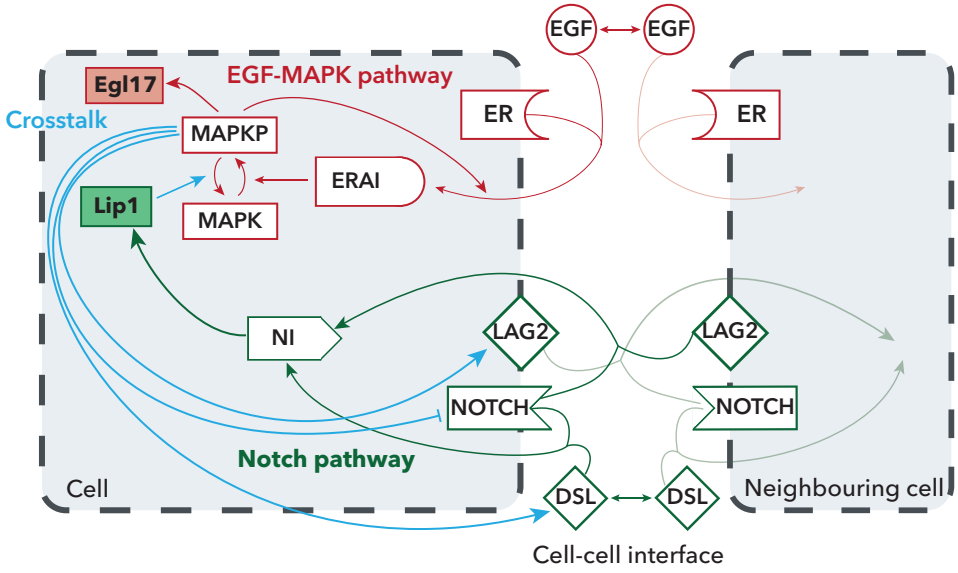


Figure 3.2: *Diagram summarizing the network of interactions involved in vulval development. Each of the dashed boxes represents a cell. EGF pathway is represented in red, Notch pathway is coloured in green and their crosstalk in cyan. All cells have the same network wiring. Boxes are gene products. Whithin the EGF pathway, EGF represents EGF ligand (lin-3); ER the EGF receptor (let-23), ERAI is the activated ER; MAPK is the MAP kinase; MAPKP is phosphorylated MAP kinase; Egl17, primary fate effector. Whithin the Notch pathway, DSL represents diffusible Delta ligand; NOTCH is the Notch receptor (lin-12); LAG2 is the membrane-bound Delta; NI represents Notch in the intracellular domain; and Lip1 is the secondary fate effector and MAP kinase phosphatase. One-way arrows indicate transformation (MAPK to MAPKP and conversely), binding (ligands to receptors), or activation of the downstream node. Two-way arrows indicate diffusion. The −| arrow from MAPKP to NOTCH indicates inhibition of the downstream node. Adapted from (Hoyos et al. 2011).*

Two models have been proposed to describe the patterning mechanism: the **morphogen model** and the **sequential model**. In the morphogen model, EGF acts as a morphogen where different levels of EGF specify the different fates by controlling different levels of MAPK pathway activity and, consequently, the fate of a cell depends on the distance to the AC (source of the morphogen EGF). On the other hand, in the sequential model, the anchor cell induces the primary fate in the closest VPC which, in turn, induces the neighbouring cells towards secondary fate through the Notch pathway; while lack of activation of these pathways would promote tertiary fate (Hoyos et al. 2011). However, there is experimental evidence

supporting both of these two mechanisms.

The development of different mutant worms have been studied by upregulating, downregulating or knocking out the expression of different genes at different times, or by ablating certain cells involved in the process. These experiments shed light on the mechanisms that regulate the development of the vulva. For example, significant reduction of EGF signal (either by AC ablation at a very early time or by EGF loss of function) produces the pattern 3°3°3°3°3° (Corson & Siggia 2012, Komatsu et al. 2008), suggesting that the activation of the EGF-Ras-MAP pathway is necessary for the induction of the VPCs. Table 3.1 contains a summary of the experimental results for some experiments that can be found in the literature (Corson & Siggia 2012).

Table 3.1: *Table of experimental data obtained from the literature (Corson & Siggia 2012, Milloz et al. 2008). Fates for P4.p and P8.p (or P5.p and P7.p) have been averaged, since we assume that the pattern is symmetrical around the anchor cell. Developmental stages from the earliest to the latest: (Let. L2) lethargic L2; (eL3) early L3; (DU) Dorsal Uterine precursor cells dividing or divided once; (VU) Ventral Uterine precursor cells dividing or divided once; (3°) 3° cells have divided; (2-cell) all Pn.p cells have divided once.*

	Experiment	VPC fates (% 1°, 2° 3°)				
		P4.p	P5.p	P6.p	P7.p	P8.p
	Wild-type outcomes under reduced signalling					
(1)	Wild type	0, 0, 100	0, 100, 0	100, 0, 0	0, 100, 0	0, 0, 100
(2)	*let-23* mosaic (no EGF receptors in P5/7.p)			wild type		
(3)	Half dose of *lin-3* (Half EGF ligand)			wild type		
(4)	Half dose of *lin-12* (Half Notch receptor)			wild type		
	Phenotypes under reduced Notch/excess EGF					
(5)	Notch null, and 2×EGF in WT (2 ACs) (No Notch receptor, twice as much WT EGF ligand)	0, 0, 100	100, 0, 0	100, 0, 0	100, 0, 0	0, 0, 100
(6)	No Notch signalling, WT EGF (No Notch receptor, WT EGF ligand)	0, 0, 100	0, 0, 100	100, 0, 0	0, 0, 100	0, 0, 100
(7)	EGF overexpression (overexpression of EGF ligand)	24,54,21	54,46,0	96,4,0	37,63,0	12,38,50
	Phenotypes following anchor cell ablation at the corresponding stages					
(8)	Let. L2	–	0, 0, 100	0, 0, 100	0, 0, 100	–
(9)	eL3	–	3,22,75	18,18,64	0,20,80	–
(10)	DU	–	0,59,41	31,38,31	0,54,46	–
(11)	VU	–	0,95,5	52,48,0	8,84,8	–
(12)	3°	–	2,98,0	65,35,0	0,100,0	–
(13)	2-cell	–	1,99,0	93,7,0	1,99,0	–

By taking advantage of the biological data and in order to deeply understand the mechanism behind this process, several mathematical and computational models have been proposed. In the following section we review the most important ones, which will help us to introduce our new mathematical model.

3.2 Previous approaches to Mathematical and Computational Modelling of Vulval Development

Different approaches for the modelling of cell specialisation in the vulva of *C. elegans* have been published. They differ in the way they abstract the process, and the *language* that they use to represent this abstraction.

Since 2005, different computational and mathematical models have been studied. Some of them use the language of statecharts to develop a diagramatic and discrete model of vulval development (Fisher et al. 2005, 2007, Kugler et al. 2008).

A more mathematical approach was taken then by Giurumescu et al. (2006) in which they tried to model a relatively simplistic version of the gene regulatory network underlying the differentiation of the VPCs by using differential equations. Hoyos et al. (2011) then presented a model for a more detailed version of the gene regulatory network (See Figure 3.2), also taking advantage of ordinary differential equations. Later on, Weinstein & Mendoza (2013), Weinstein et al. (2015) studied a very complex network believed to regulate vulval development in *C. elegans*, but they took advantage of Boolean networks instead.

More recent studies have developed a more qualitative approach, going away from these very complex networks and sumarising the essence of the process, proving to be a very successful way of understanding the process (Corson & Siggia 2012, Ping & Tang 2015, Corson & Siggia 2017).

In particular, Corson & Siggia (2012, 2017) develop a very elegant model inspired by the idea of *Waddington's epigenetic landscapes* (Waddington 1957), as we described in Section 2.3.

The landscape that Corson & Siggia (2012, 2017) propose takes values in a two-dimensional space ($\mathbb{R}^2$). They fix the topology of the model to be derived from a three-fold symmetric cubic potential (Equation 3.1), so that it has three attractors (one for each fate) and all transitions between two fates are permissible.

$$V(x, y) = -x^2 - y^2 + x^2 y - \frac{1}{3}y^3. \tag{3.1}$$

The flow derived from that potential is given by:

$$\vec{f}(x, y) = \left(-\frac{\partial V(x, y)}{\partial x}, -\frac{\partial V(x, y)}{\partial y}\right) = (2x(1 - y), 2y + (y^2 - x^2)). \tag{3.2}$$

The state of each cell in time is represented by the solution of an ordinary differential

equation derived from such potential. They describe the state of each cell by a two-dimensional vector $\vec{r}(t) = (x(t), y(t))$ at time t and the change of position in time is given by the equation

$$\frac{d\vec{r}}{dt} = [\vec{\sigma}_1(\vec{f}) - \vec{r}], \tag{3.3}$$

where

$$\vec{\sigma}_1(\vec{f}) = \tanh(\|\vec{f}\|)\frac{\vec{f}}{\|\vec{f}\|},$$

which has three attractors at $(x_1^*, y_1^*) = \left(\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$, $(x_2^*, y_2^*) = (0, 1)$ and $(x_3^*, y_3^*) = \left(-\frac{\sqrt{3}}{2}, -\frac{1}{2}\right)$, representing fates 1°, 2° and 3° respectivelly.

Since the landscape needs to be dynamic, i.e. different attractors are accessible at different points in time (which will be related to the signals that the VPC is receiving), they add some parameters to Equation 3.4:

$$\frac{d\vec{r}}{dt} = [\vec{\sigma}_1(\vec{f} + \vec{m}) - \vec{r}], \tag{3.4}$$

where $\vec{m} = \vec{m}_0 + \vec{m}_1 + \vec{m}_2$, and $\vec{m}_0 = m_0(x_3^*, y_3^*)$, $\vec{m}_1 = m_1(x_1^*, y_1^*)$ $\vec{m}_2 = m_2(x_2^*, y_2^*)$. This way, the "strength" of the attractors will be related to the modulus of the corresponding vectors $\vec{m}_i$.

Since EGF signal promotes the 1° fate, $\vec{m}_1$ represents the EGF signal that the VPC receives. Similarly, $\vec{m}_2$ represents the response of the VPC to Notch signaling and it promotes the fate 2°. On the other hand, $\vec{m}_0$ biases the flow towards the default 3° state.

Finally, they add some white noise to the system given in Equation 3.4 in order to account for intrinsic biological noise. This model allows them to fit a big set of experimental data (that were not reproduced by previous models), and to give very interesting predictions.

We think that this is a very powerful approach, but there is some space for improvement.

In their model, the vector $\vec{m}$ represents the *control parameters* of the system. However, due to the form of the dynamical system, it is very difficult to study the stability of the system depending on the values of these parameters, and impossible to give an analytical expression of the bifurcation set, which would give very useful information about the model. With the bifurcation set we can directly map parameter values to number of attractors very easily.

We also think that they make an adhoc choice of the topology for the system. The choice of such topology is the most important part of the model, and it would

be useful to have a method to tailor different topologies to different processes.

These are the reasons why, in this chapter, we take advantage of catastrophe theory to develop a new model for vulval development. In particular, we aim to find a simple dynamical system that describes the process of differentiation of the VPCs. This dynamical system will be derived from two of the elementary catastrophes in Thom's Classification Theorem (See Theorem 2 Chapter 2), and will allow us to see the process from a new perspective.

3.3 A new mathematical approach

Inspired by the models by Corson & Siggia (2012, 2017), we introduce an innovative and more natural approach to the modelling of vulval development by taking advantage of catastrophe theory. Thom's classification theorem (Theorem 2 in Chapter 2) allows us to classify a rich class of dynamical systems and the bifurcations between them. The elementary chatastrophes are polynomial equations in at most two variables and a few control parameters. The idea is that we can use these relatively simple dynamical systems as "LEGO blocks" to build a more complex one with the desired characteristics, which dynamics might not be reproducible by a polynomial equation.

In this section we will first introduce the topology chosen for the model; secondly, we will show how to write the mathematical control parameters in terms of the biological parameters or signals that control the process; we will then explain the model proposed and, finally, we will show how to implement it and simulate data from it.

3.3.1 Choice of the landscape's topology

The gene regulatory network described in Figure 3.2 suggests how the topology of the space should be.

In contrast to Corson & Siggia (2012, 2017), we believe that not all the fates are equivalent, but that the tertiary fate is a *special* fate. It is the default state for the cell if it does not receive any signal; and it is also a state to which the cell does not go back once it is partially differentiated, even if the signals are switched off (as the anchor cell ablation experiments show (Milloz et al. 2008)). Transitions from fate 3° to fate 1°, and fate 3° to fate 2°, are allowed as Sternberg & Horvitz (1989) suggests. Also, the transition from fate 2° to fate 1° is achieved by increasing EGF signal and turning Notch signal off (Greenwald et al. 1983). These are the reasons

why we think that the topology of decisions between the three fates should be the one given in Figure 3.3.

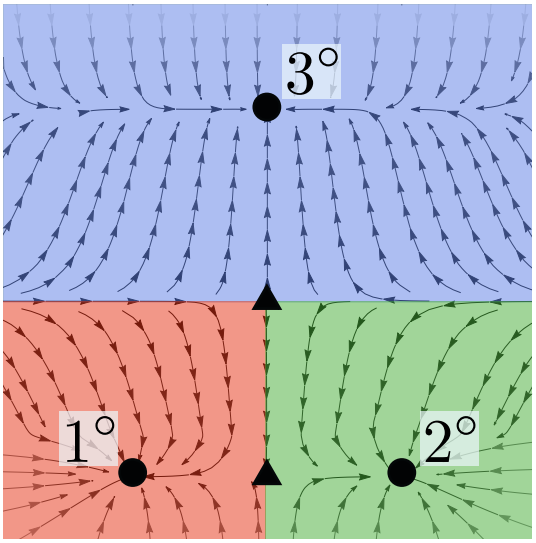


Figure 3.3: *Topology chosen for the mathematical model of vulval development proposed. Minima or attractors are represented by black circles, and saddles are represented by black triangles. Attractors represent* $1°$*,* $2°$ *and* $3°$ *fates respectively. Basins of attraction are coloured red, green and blue for* $1°$*,* $2°$ *and* $3°$ *fates respectively. The saddle between* $3°$ *and* $1° - 2°$ *regions will represent a binary decision between* $3°$ *and vulva. The saddle between* $1°$ *and* $2°$ *regions represents a decision between taking* $1°$ *or* $2°$ *fate.*

In order to find a system that contains this topology, we will take advantage of catastrophe theory. As we saw in Chapter 2, the catastrophe manifold and bifurcation set help us to have a map between state and parameter spaces.

With the purpose of defining the new model, we need to first introduce the following dynamical system:

$$\begin{cases} \frac{dx}{dt} &= \mathbb{1}_{\mathbb{R}_{\leq 0}}(y) f(x,a,b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(y))x = f_1(x,y,a,b,c,M) \\ \frac{dy}{dt} &= -y((y-M)^2 + c) = f_2(x,y,a,b,c,M), \end{cases} \tag{3.5}$$

where $f(x,a,b) = -4x^3 - 2ax - b$, flow corresponding to the cusp catastrophe discussed in Chapter 2, $\mathbb{1}_{\mathbb{R}_{\leq 0}}(y)$ is the indicator function in $[-\infty, 0)$, i.e.

$$\mathbb{1}_{\mathbb{R}_{\leq 0}}(y) = \begin{cases} 0 & \text{if} \quad y > 0 \\ 1 & \text{if} \quad y \leq 0 \end{cases} \tag{3.6}$$

and $a, b, c, M \in \mathbb{R}$ are parameters. Let us study the stability of the system with respect to the parameters a, b and c and $M > 0$.

3.3.2 Stability of the dynamical system

The critical points of this system of differential equations are given by the zeros of $\frac{dx}{dt}$ and $\frac{dy}{dt}$ in Equation 3.5.

First, we focus on the zeros of the equation regarding the change in time of y, since it does not involve x. Suppose that we were interested in knowing the stability of the following one dimensional system on y with $M > 0$ and c as parameters:

$$\frac{dy}{dt} = -y((y - M)^2 + c) \tag{3.7}$$

The possible zeros of $\frac{dy}{dt}$ are $y_1^* = M + \sqrt{-c}$, $y_2^* = M - \sqrt{-c}$ and $y_3^* = 0$. Since we assume that the state variables are real, y_1^* or y_2^* will only exist if $c \leq 0$.

The stability of the points y_i^* will depend on the sign of $\frac{d}{dy}\left(\frac{dy}{dt}\right)$ evaluated at the corresponding equilibrium point. Fixing the parameter $M > 0$, Figure 3.4 shows the corresponding bifurcation diagram.

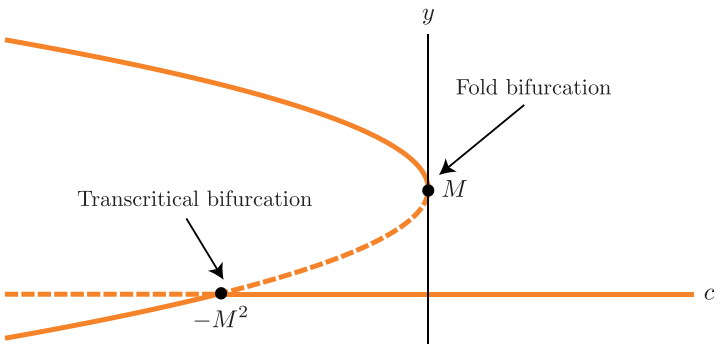


Figure 3.4: *Bifurcation diagram of Equation* (3.7). *With $M > 0$ fixed, the orange lines show the equilibria of the system for different values of c. Continuous lines represent stable equilibria and dashed lines represent unstable equilibria.*

We are only interested in the fold bifurcation that happens when $c = 0$, and not in the transcritical bifurcation at $c = -M^2$. Therefore we will suppose that

$$c > -M^2 \tag{3.8}$$

from this point onwards. Summing up, depending on the value of c, Equation (3.7) has one ($c > 0$), two ($c = 0$) or three ($c < 0$) equilibria.

Going back to the two dimensional system of equations in 3.5, this means that the system will have equilibrium points with one, two or three possible different values of the y-coordinate, depending on the value of the parameter c.

Now let us consider the system described in Equation (3.5). In order to

find the equilibrium points of the whole system we need to find the zeros of both equations. For $\frac{dy}{dt} = 0$, as we saw before, the equation will have one, two or three zeros depending on the value of c. Let us now study the zeros of $\frac{dx}{dt}$ in the three cases separately:

1. **Negative values of** c $(c > -M^2)$

   In this case, $\frac{dy}{dt} = 0$ has three possible roots: $y_1^* = M + \sqrt{-c}$, $y_2^* = M - \sqrt{-c}$ and $y_3^* = 0$.

   If $y = y_1^* = M + \sqrt{-c}$, since $\mathbb{1}_{\mathbb{R}_{\leq 0}}(M + \sqrt{-c}) = 0$,

   $$\frac{dx}{dt} = \mathbb{1}_{\mathbb{R}_{\leq 0}}(M + \sqrt{-c})f(x, a, b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(M + \sqrt{-c}))x = -x = 0 \Leftrightarrow x = 0.$$

   Therefore there is an equilibrium point at $(0, M + \sqrt{-c})$.

   Also, if $y = y_2^* = M - \sqrt{-c}$, since $\mathbb{1}_{\mathbb{R}_{\leq 0}}(M - \sqrt{-c}) = 0$,

   $$\frac{dx}{dt} = \mathbb{1}_{\mathbb{R}_{\leq 0}}(M - \sqrt{-c})f(x, a, b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(M - \sqrt{-c}))x = -x = 0 \Leftrightarrow x = 0.$$

   Hence, there is another equilibrium point at $(0, M - \sqrt{-c})$.

   And if $y = y_3^* = 0$, since $\mathbb{1}_{\mathbb{R}_{\leq 0}}(0) = 1$,

   $$\frac{dx}{dt} = \mathbb{1}_{\mathbb{R}_{\leq 0}}(0)f(x, a, b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(0))x = f(x, a, b) = 0.$$

   Therefore the values for the x-coordinate correspond to the equilibria of a cusp catastrophe with parameters a and b. Depending on the value of the discriminant $\Delta = 8a^3 + 27b^2$, the equation will have one ($\Delta > 0$), two ($\Delta = 0$) or three ($\Delta < 0$) real roots. Therefore, depending on the value of Δ, there will be one, two or three equilibria on the x-axis.

   The stability of the points can be checked by looking at the eigenvalues of the jacobian matrix:

   $$J_{(a,b,c)}(x^*, y^*) = \left(\begin{array}{cc} \frac{\partial f_1(x,y,a,b,c,M)}{\partial x} & \frac{\partial f_1(x,y,a,b,c,M)}{\partial y} \\ \frac{\partial f_2(x,y,a,b,c,M)}{\partial x} & \frac{\partial f_2(x,y,a,b,c,M)}{\partial y} \end{array} \right) \Bigg|_{(x^*,y^*)} \tag{3.9}$$

   where (x^*, y^*) is an equilibrium of the system in Equation 3.5, with parameters a, b, c, M. Since $\frac{\partial f_2(x,y,a,b,c,M)}{\partial x} = 0$, for all values of x, y, a, b, c, M, the jacobian matrix is upper triangular and the eigenvalues are given by the elements on the diagonal. Figure 3.5 shows the possible configurations of the equilibrium

points in the state space.

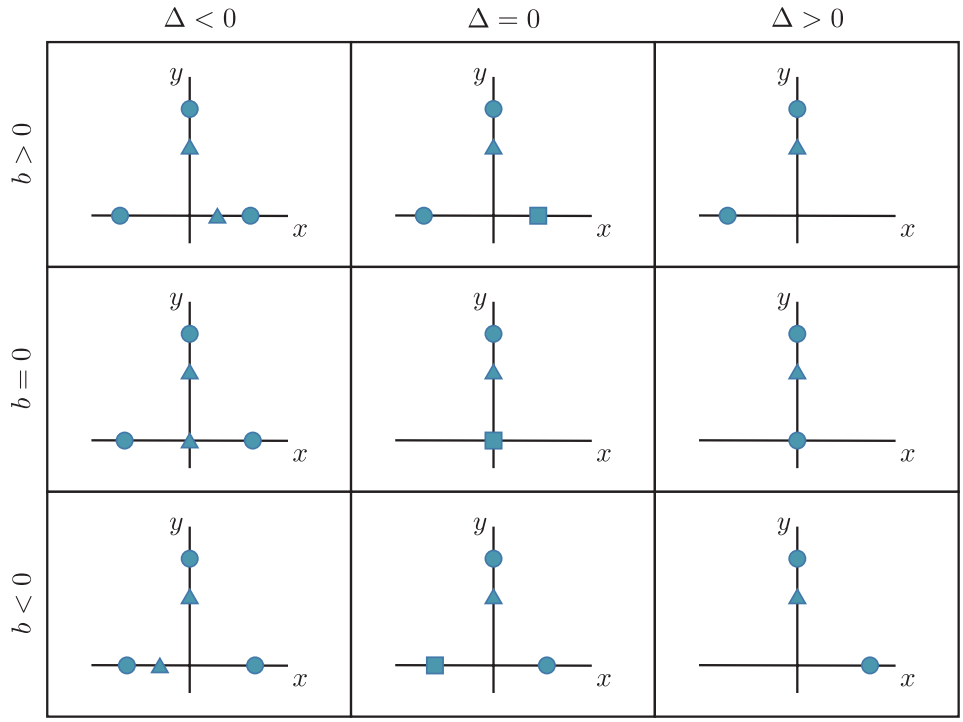


Figure 3.5: *Sketch of the equilibria of the system for $-M^2 < c < 0$ depending on the values of b and Δ. Stable equilibria are represented by circles. Saddles are represented by triangles. Degenerate steady states are represented by squares.*

2. c **equal to zero**

The roots of $-y((y-M)^2+c)=0$ are $y_1^* = M$ and $y_2^* = 0$.

If $y = y_1^* = M$, since $\mathbb{1}_{\mathbb{R}_{\leq 0}}(M) = 0$,

$$\frac{dx}{dt} = \mathbb{1}_{\mathbb{R}_{\leq 0}}(M) f(x,a,b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(M))x = -x = 0 \Leftrightarrow x = 0.$$

And if $y = y_2^* = 0$, since $\mathbb{1}_{\mathbb{R}_{\leq 0}}(0) = 1$,

$$\frac{dx}{dt} = \mathbb{1}_{\mathbb{R}_{\leq 0}}(0) f(x,a,b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(0))x = f(x,a,b) = 0.$$

Therefore the system has one equilibrium with x-coordinate equal to 0 (the point $(x_1^*, y_1^*) = (0, M)$). Indeed, this is a fold bifurcation point and this equilibrium point is a degenerate point. As before, depending on the value of $\Delta = 8a^3 + 27b^2$, there will also be one ($\Delta > 0$), two ($\Delta = 0$) or three ($\Delta < 0$) equilibria on the x-axis. Again, their stability can be checked by evaluating

the jacobian in Equation (3.9) at the points of interest.

Figure 3.6 shows the possible configurations in the state space.

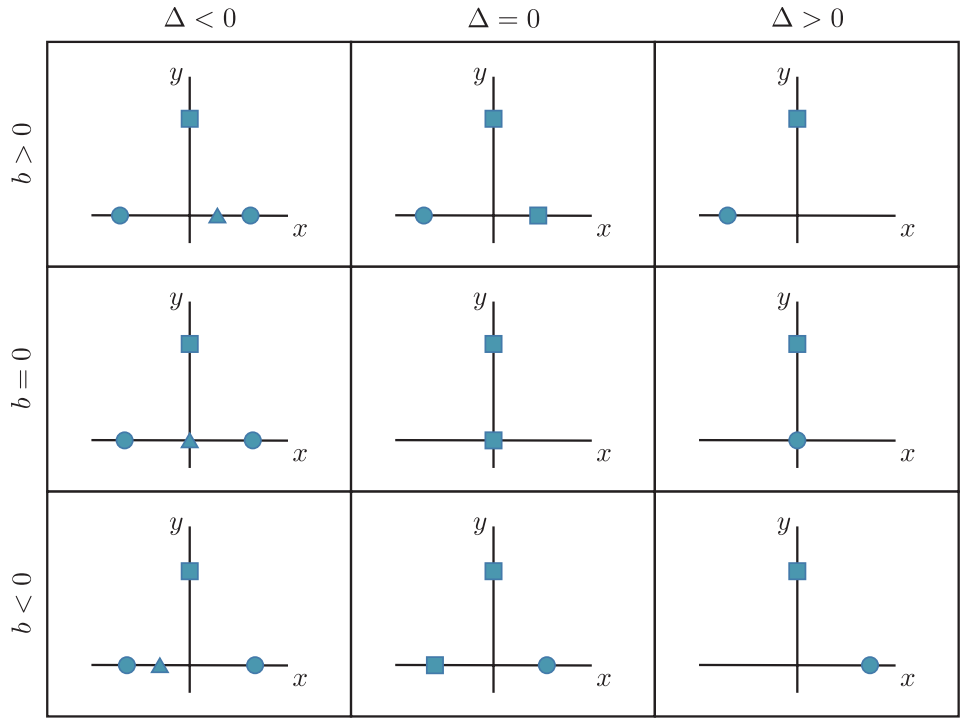


Figure 3.6: *Sketch of the equilibria of the system for $c = 0$ depending on the values of b and Δ. Stable equilibria are represented by circles. Saddles are represented by triangles. Degenerate steady states are represented by squares.*

3. **Positive values of c**

   The only root of $-y((y-M)^2+c) = 0$ is $y_1^* = 0$.

   Substituting in $\frac{dx}{dt} = 0$ we have that

   $$\mathbb{1}_{\mathbb{R}_{\leq 0}}(0) f(x,a,b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(0))x = f(x,a,b) = -4x^3 - 2ax - b = 0$$

   thus, as on the previous cases, depending on the value of $\Delta = 8a^3 + 27b^2$, the system will have one ($\Delta > 0$), two ($\Delta = 0$) or three ($\Delta < 0$) equilibria on the x-axis.

   The configurations of the equilibria in the state space, for this case, depending on the value of Δ can be seen in Figure 3.7.

Summing up, the system can have 5 to 1 equilibrium points in the state space depending on the values of the parameters a, b, c. There is a fold bifurcation point

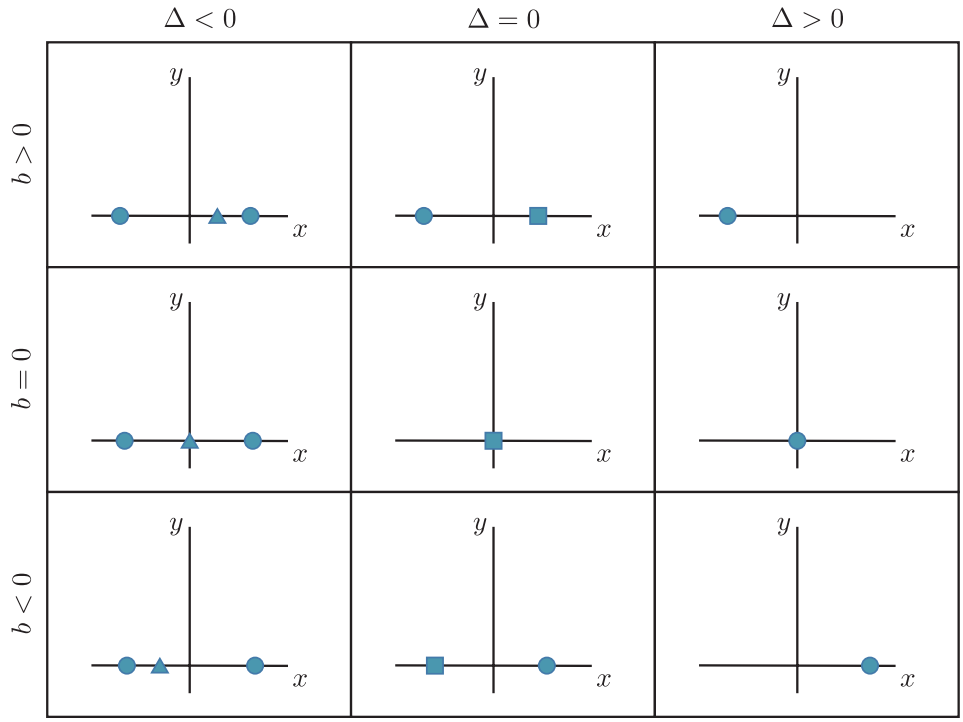


Figure 3.7: *Sketch of the equilibria of the system for $c > 0$ depending on the values of b and Δ. Stable equilibria are represented by circles. Saddles are represented by triangles. Degenerate steady states are represented by squares.*

on the y-axis controlled by the parameter c, and a cusp catastrophe on the x-axis controlled by the parameters a and b.

On a side note, the value of the parameter M does not really change the stability of the system, qualitatively. It only affects the interval in which c can take values, and the coordinates of the equilibrium point. We have decided to take positive values of M, and $\mathbb{1}_{\mathbb{R}_{\leq 0}}(y)$ to be the step function which value is 1 for $y \in (-\infty, 0)$. If M was negative, we would need to take a different step function which value is 1 for $y \in [0, \infty)$. In this case, the equilibria would have y-coordinates with the same absolute value as the ones before, but with the opposite sign (the bifurcation set in Figure 3.4 would be a reflection of the original one with respect to the c-axis), and the configurations given in Figures 3.5,3.6 and 3.7 would be reflections of the originals one with respect to the x-axis. Therefore, we only consider a, b and c as control parameters.

We can now couple each fate to an attractor. We assign 3° fate to the equilibrium point $(0, M+\sqrt{-c})$, when it exists; and the stable equilibria of the form $(x^*, 0)$ will correspond to 1° fate if $x^* < 0$ and to 2° fate if $x^* > 0$ (see Figure 3.8). The binary decision between 3° and vulval fates is represented by a fold bifurcation; while the binary decision between taking 1° or 2° fate is represented by a cusp bifurcation.

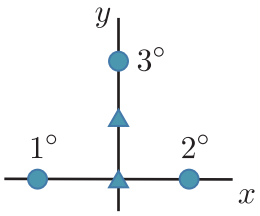


Figure 3.8: *Example of fate assignment for each attractor when* $c < 0, \Delta < 0, b = 0$.

3.3.3 Catastrophe manifold of the system

We can also study the stability of the system by taking advantage of catastrophe theory. As defined in Chapter 2, we call the **catastrophe manifold**, $\mathcal{M}$, the set of points $(x, y, a, b, c) \in \mathbb{R}^5$ such that (x, y) is an equilibrium point for the system in Equation (3.5) for parameter values a, b and c. In other words,

$$\mathcal{M} = \{(x, y, a, b, c) \in \mathbb{R}^5 : f_1(x, y, a, b, c) = f_2(x, y, a, b, c) = 0\}.$$

Here we will show that $\mathcal{M}$ is a manifold that can be written as the union of two sets, one that corresponds to the fold bifurcation and another one that corresponds to the cusp bifurcation. Indeed, $(x, y, a, b, c) \in \mathcal{M}$ if and only if

$$f_1(x, y, a, b, c) = \mathbb{1}_{\mathbb{R}_{\leq 0}}(y)(-4x^3 - 2ax - b) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(y))x = 0 \tag{3.10a}$$

and

$$f_2(x, y, a, b, c) = -y((y - M)^2 + c) = 0. \tag{3.10b}$$

From Equation (3.10b) we have that $y = 0$ or $y = M \pm \sqrt{-c}$. Now, if $y = 0$, from Equation (3.10a):

$$-4x^3 - 2ax - b = 0 \iff b = -4x^3 - 2ax.$$

And if $y = M \pm \sqrt{-c}$, from Equation (3.10a), x must be 0.

We can then express $\mathcal{M}$ as the disjoint union of two three-dimensional man-

ifolds, $\mathcal{M}_1$ and $\mathcal{M}_2$, where:

$$\mathcal{M}_1 = \{(x, 0, a, -4x^3 - 2ax, c) \in \mathbb{R}^5 : x, a, c \in \mathbb{R}\} \tag{3.11a}$$

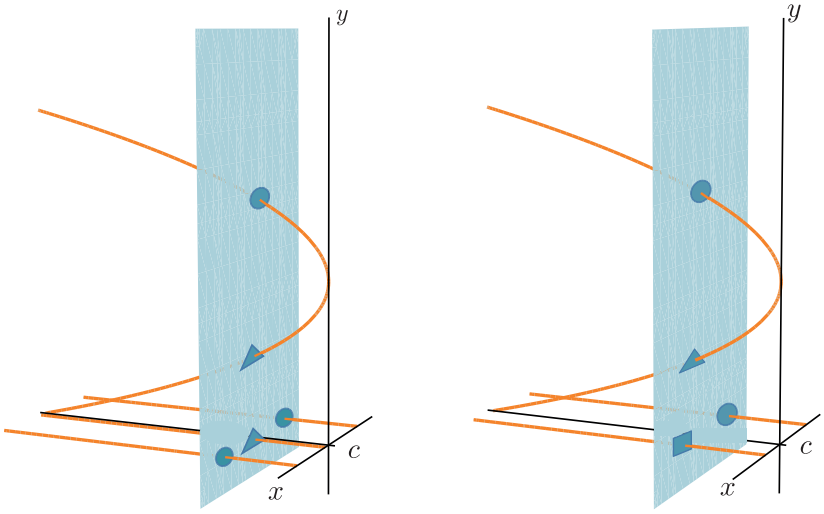
$$\mathcal{M}_2 = \{(0, y, a, b, -(y - M)^2) \in \mathbb{R}^5 : a, b \in \mathbb{R}, y \in (0, 2M)\} \tag{3.11b}$$

Therefore $\mathcal{M} = \mathcal{M}_1 \cup \mathcal{M}_2$ is a three-dimensional manifold in $\mathbb{R}^5$.

In order to visualise $\mathcal{M}$ we can plot the intersection with the three-dimensional space

$$\mathcal{N}_{(a_0,b_0)} = \{(x, y, a, b, c) \in \mathbb{R}^5 : a = a_0, b = b_0\}$$

to see how the equilibrium points change when c changes (see Figure 3.9).



(A) $a_0 = -1, b_0 = 0.$ **(B)** $a_0 = -1.5, b_0 = 1.$

Figure 3.9: *Visualisation of the catastrophe manifold of the system in Equation 3.5.In orange, $\mathcal{M} \cap \mathcal{N}_{(a_0,b_0)}$ for different values of a_0, b_0. The plane π in cyan is obtained by fixing $c = c_0 < 0$, in other words, $\pi = \{(x, y, c) \in \mathbb{R}^3 : c = c_0\}$. The intersection of π and $\mathcal{M} \cap \mathcal{N}_{(a_0,b_0)}$ gives the equilibria of the system for parameters $a = a_0, b = b_0, c = c_0$. If sliding the plane π along the c axis, the equilibria on the y axis will move along the parabola. As before, circles represent stable equilibria, triangles represent saddles and squares represent degenerate points.*

3.3.4 Singularity and bifurcation sets of the system

The **bifurcation set**, as defined in Chapter 2, lets us characterise regions in the control space that will give rise to certain configurations of the equilibria in the state space. By understanding the bifurcation set of the system, we can predict which equilibrium points will be present in the state space when knowing the values of the

control parameters.

We first find an expression for the **singularity set** of the system, $\mathcal{S}$, defined in Chapter 2, which is given by:

$$\mathcal{S} = \mathcal{M} \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : det(J_{(a,b,c)}(x, y)) = 0\}, \tag{3.12}$$

which is the subset of degenerate points of $\mathcal{M}$ (see Chapter 2).

Let us denote $\frac{\partial f_1}{\partial x}$ by $D_x f_1$ and $\frac{\partial f_2}{\partial y}$ by $D_y f_2$. Since $det(J_{(a,b,c)}(x, y)) = D_x f_1 D_y f_2$, as explained in Subsection 3.3.2,

$$det(J_{(a,b,c)}(x, y)) = 0 \iff \begin{cases} D_x f_1(x, y, a, b, c) = \mathbb{1}_{\mathbb{R}_{\leq 0}}(y)(-12x^2 - 2a) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(x)) = 0 \\ \quad \text{or} \\ D_y f_2(x, y, a, b, c) = -((y - M)^2 + c) - 2y(y - M) = 0 \end{cases}$$

Therefore we can write $\mathcal{S} = \mathcal{S}_1 \cup \mathcal{S}_2$ where

$$\mathcal{S}_1 = \mathcal{M} \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_x f_1(x, y, a, b, c) = 0\} \tag{3.13}$$

and

$$\mathcal{S}_2 = \mathcal{M} \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_y f_2(x, y, a, b, c) = 0\}. \tag{3.14}$$

Let us understand the geometry of these sets. The set $\mathcal{S}_1$ can be written as

$$\mathcal{S}_1 = (\mathcal{M}_1 \cup \mathcal{M}_2) \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_x f_1(x, y, a, b, c) = 0\},$$

where $\mathcal{M}_1$ and $\mathcal{M}_2$ are defined in Equations 3.11a and 3.11b.

Applying the distributive property of the intersection over the union,

$$\mathcal{M}_1 \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_x f_1(x, y, a, b, c) = 0\} = \{(x, 0, -6x^2, 8x^3, c) : x, y, c \in \mathbb{R}\},$$

and

$$\mathcal{M}_2 \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_x f_1(x, y, a, b, c) = 0\} = \varnothing.$$

Therefore,

$$\mathcal{S}_1 = \{(x, 0, -6x^2, 8x^3, c) : x, y, c \in \mathbb{R}\}, \tag{3.15}$$

which corresponds to a cusp for every value of c (See Equation 2.36 in Chapter 2).

On the other hand, the set $\mathcal{S}_2$ can be written as

$$\mathcal{S}_2 = (\mathcal{M}_1 \cup \mathcal{M}_2) \cap \{(x, y, a, b, c) \in \mathbb{R}^5 : D_y f_2(x, y, a, b, c) = 0\}.$$

Again, applying the distributive property of the intersection over the union,

$$\mathcal{M}_1 \cap \{(x,y,a,b,c) \in \mathbb{R}^5 : D_y f_2(x,y,a,b,c) = 0\} = \{(x,0,a,-4x^3-2ax,-M^2) : a,c \in \mathbb{R}\}.$$

But, since we are only taking into account values of $c > -M^2$, we don't consider this set.

Moreover,

$$\begin{aligned}\mathcal{M}_2 \cap \{(x,y,a,b,c) \in \mathbb{R}^5 : D_y f_2(x,y,a,b,c) = 0\} = \\ = \{(0,0,a,b,-M^2), a,b \in \mathbb{R}\} \cup \{(0,M,a,b,0), a,b \in \mathbb{R}\},\end{aligned}$$

where, again, we do not consider the first subset since $c > -M^2$. Therefore

$$\mathcal{S}_2 = \{(0,M,a,b,0), a,b \in \mathbb{R}\}. \tag{3.16}$$

From Equations (3.15) and (3.16) we can write the singularity set, $\mathcal{S}$, of the system as:

$$\mathcal{S} = \{(x,0,-6x^2,8x^3,c) : x,y,c \in \mathbb{R}\} \cup \{(0,M,a,b,0), a,b \in \mathbb{R}\} \tag{3.17}$$

Finally, projecting $\mathcal{S}$ into the control space we obtain an expression for the bifurcation set, $\mathcal{B}$ (See Chapter 2 for a definition):

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{(a,b,c) \in \mathbb{R}^3 : 8a^3 + 27b^2 = 0\} \cup \{(a,b,0) : a,b \in \mathbb{R}\} \tag{3.18}$$

which is a two-dimensional set in the three-dimensional control space (See Figure 3.10).

Figures 3.7, 3.6 and 3.5 give a description of the different landscapes in the state space depending on the values of the control parameters (a,b,c). We can see that crossing the orange plane in Figure 3.10 would bifurcate the points $(0, M-\sqrt{-c})$ and $(0, M+\sqrt{-c})$ and crossing the purple surface in Figure 3.10 would bifurcate the points on the x-axis.

We have obtained a map between the parameter space and the state space. Given a parameter value we know the topology of the state space. Continuing with the simile, we know how to change the parameters to make certain drains or attractors stronger or weaker or, even, to disappear. This gives us a description of how the topology of the space changes as we vary the parameters and therefore helps us to define the dynamical system that will describe the state of each VPC.

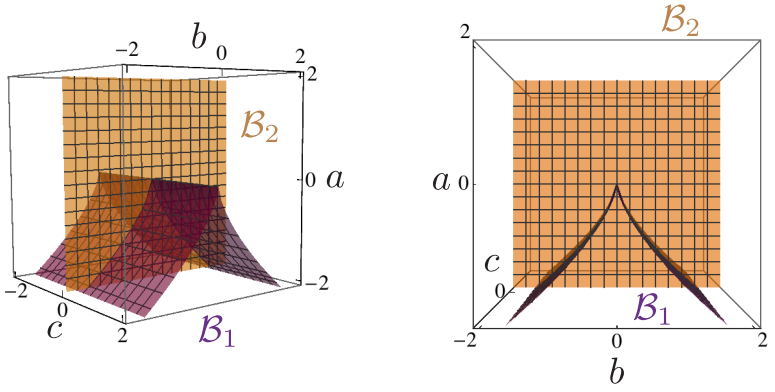


Figure 3.10: *Plots showing different views of the bifurcation set given in Equation* (3.18). *The set* $\mathcal{B}_1$ *corresponds to the purple cuspoidal cylinder, while the set* $\mathcal{B}_2$ *corresponds to the orange plane. For every value of* c *the system has a cusp bifurcation set with* a, b *as parameters.Also, for any value of* a, b*, the points* $(0, M - \sqrt{-c})$ *and* $(0, M + \sqrt{-c})$ *bifurcate when* $c = 0$*, hence the plane.*

.

3.3.5 Change of parameters

In the previous section we have defined a dynamical system with a two-dimensional state space and a three-dimensional control space, and we have classified the possible landscapes that can appear in the state space. The goal is to take advantage of this dynamical system to define a system of differential equations such that its solution is a description of the state of each VPC during the process of differentiation.

As described in Section 3.1, this process is controlled by two signalling pathways: induction from the anchor cell by EGF signal and lateral signalling through Notch. In consequence, one can think that the dynamics of the real system are controlled by these two parameters: level of EGF that the cell is receiving (this parameter will be denoted by EGF) and level of Notch that the cell is receiving (parameter that will be denoted by $NOTCH$). This means that the topology of the state space in the mathematical model should be related to the parameters EGF and $NOTCH$. In other words, the parameters a, b, c in Equation (3.5) should be functions of EGF and $NOTCH$.

Once one finds a suitable way of writing a, b, c as functions of EGF and $NOTCH$, the state of a vulval precursor cell will be described by the solution of

the system of ordinary differential equations:

$$\begin{cases} \frac{dx}{dt} &= \mathbb{1}_{\mathbb{R}_{\leq 0}}(y) f(x, a(EGF, NOTCH), b(EGF, NOTCH)) - (1 - \mathbb{1}_{\mathbb{R}_{\leq 0}}(y))x \\ \frac{dy}{dt} &= -y((y - M)^2 + c(EGF, NOTCH)), \end{cases} \tag{3.19}$$

where f and $\mathbb{1}_{\mathbb{R}_{\leq 0}}$ were described in Section 3.3 and EGF and $NOTCH$ are the levels of EGF and Notch signals that the VPC receives in time.

Now, assuming that the functions $a(EGF, NOTCH)$, $b(EGF, NOTCH)$ and $c(EGF, NOTCH)$ are found, the corresponding values of the control parameters that describe the attractors of the system can be found for each value of EGF and $NOTCH$. Therefore, for each value of the signals, the landscape of the state space is known.

EGF and $NOTCH$ can be considered a coordinate basis that generate a two-dimensional space. Let us call it the **signal space**. The coordinates of a point in that space will represent the values of the EGF and Notch signal that a cell receives.

Our goal is to find a correspondence between the signal space and the state space. That is to say that we aim to find which topology in the state space will correspond to which point in the signal space. And we can achieve this by finding a transformation from the $\{EGF, NOTCH\}$ coordinate system to the $\{a, b, c\}$ one which, in turns, is mapped to the state space by the dynamical system defined in 3.19 (see Figure 3.11).

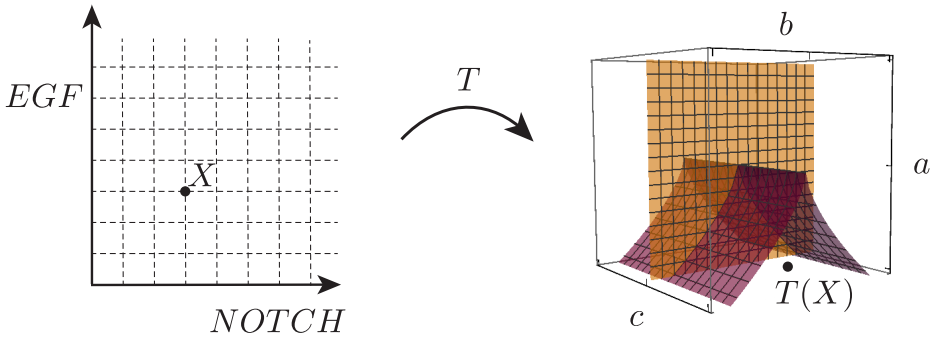


Figure 3.11: *Change of coordinates from the signal space to the control space. A transformation T will map a point X in the signal space (on the left) to a point $T(X)$ in the control space (on the right).*

For simplicity, we will focus on affine transformations that map the signal space into the control space. This affine transformation, T, will be an embedding from the affine space $\mathbb{R}^2$ to the affine space $\mathbb{R}^3$, that maps the signal space into

a plane in $\mathbb{R}^3$. These transformations will be characterised by the intersection of the planes that they define and the bifurcation set $\mathcal{B}$, since that will determine the possible dynamics of the system.

These affine transformations will have the form:

$$T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3$$
$$X = (EGF, NOTCH) \mapsto T(X) = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{pmatrix} \begin{pmatrix} EGF \\ NOTCH \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \tag{3.20}$$

where m_{ij} and q_i define the transformation. This transformation maps a point in the signal space to a point in the plane:

$$\pi_T = \{(a, b, c) \in \mathbb{R}^3 : Aa + Bb + Cc = D\} \tag{3.21}$$

in the control space where

$$A = m_{31}m_{22} - m_{21}m_{32} \tag{3.22}$$

$$B = m_{11}m_{32} - m_{31}m_{12} \tag{3.23}$$

$$C = m_{12}m_{21} - m_{11}m_{22} \tag{3.24}$$

$$D = -Aq_1 - Bq_2 - Cq_3 \tag{3.25}$$

This plane intersects the bifurcation set $\mathcal{B}$ in different subsets depending on the values of m_{ij}, and the origin of the $\{EGF, NOTCH\}$ coordinate space in that plane will be determined by the parameters q_i (See Figure 3.12).

We will assume that

$$m_{11}^2 + m_{21}^2 + m_{31}^2 = 1, \tag{3.26}$$

$$m_{12}^2 + m_{22}^2 + m_{32}^2 = 1, \tag{3.27}$$

$$m_{11}m_{12} + m_{21}m_{22} + m_{31}m_{32} = 0, \tag{3.28}$$

i.e. T is conformal (it preserves angles).

We also assume that the origin of the $\{EGF, NOTCH\}$ coordinate system is mapped to a point in which there is tristability, this is

$$\begin{cases} 8q_1^3 + 27q_2^2 < 0 \\ q_3 < 0. \end{cases} \tag{3.29}$$

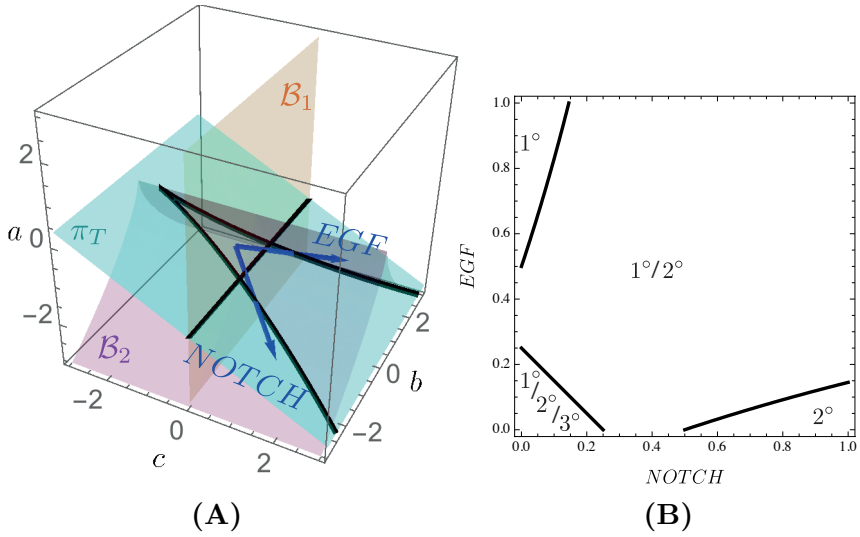


Figure 3.12: *An example of linear transformation T from the signal space to the control space. (A) The figure shows, in the control space, the plane π_T in light blue, the bifurcation set $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ in orange and purple, the intersection $\pi_T \cap \mathcal{B}$ in thick black lines and the transformation of the $(NOTCH, EGF)$ axis in dark blue. For simplicity, we have named this new coordinate system in the control space as the original in the signal space. (B) Signal space. The thick black lines correspond to the set $T^{-1}(\pi_T \cap (\mathcal{B}_1 \cup \mathcal{B}_2))$. These lines define regions in the signal space that will generate different configurations of attractors in the state space. In each region, the fates which attractors will be present for a value of signals in the corresponding region are written. For $EGF = 0 = NOTCH$, the system will contain the three attractors (one for each fate). For a medium value for EGF and $NOTCH$ only the attractors corresponding to $1°$ and $2°$ fates will be present. For high values for EGF and low values of $NOTCH$ only the attractor corresponding to $1°$ fate will be present. And, for low values for EGF and high values of $NOTCH$, only the attractor corresponding to $2°$ fate will be present.*

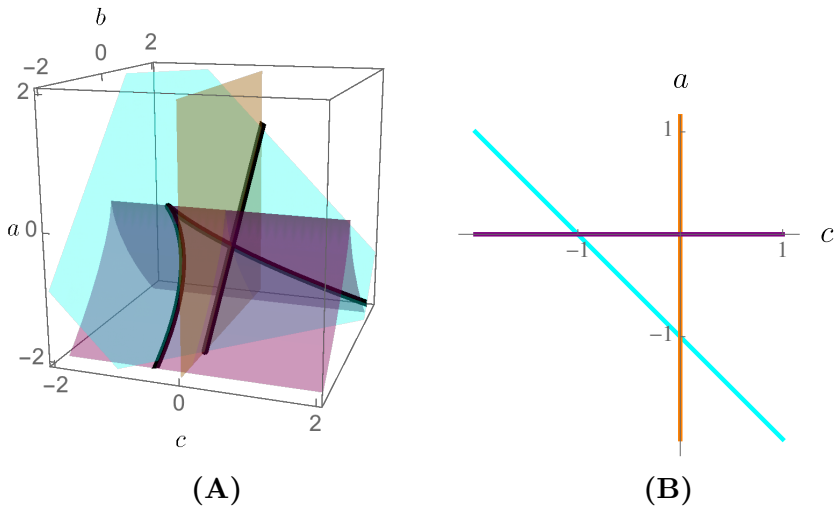
This is because of the definition of commitment of stem cells. Once a stem cell is commited, signals can be removed and the stem cell will not alter its fate. In the process of differentiation of the VPCs of a WT worm, they become specified into the three different fates (See Table 3.1) (Sternberg 2005). If we imagine the signal history of a VPC as a path in the signal space, this means that after having received different signals during their development, if the signals are switched off and the path returns to the point $EGF = NOTCH = 0$, the three attractors must be present for each VPC to be able to specify to their corresponding WT fates (P4.p specifies into $3°$, P5.p specifies into $2°$ and P6.p specifies into $1°$).

We could constrain the transformation so that high EGF pushes towards monostability of $1°$ fate (i.e. $m_{21} > 0$) and high $NOTCH$ pushes towards monostability of $2°$ fate (i.e. $m_{22} < 0$). However, we decide not to constraint the system so much, and see what the data fitting decides as the best strategy.

Finally, considering possible intersections between the plane π_T and the bifurcation set (see Appendix B for a classification), we decide to only accept transformations such that $A \neq 0 \neq C$. This is not an important constraint, since the sets $A = 0$ or $C = 0$ in $\mathbb{R}^3$ have zero measure. Moreover, by taking $A = \varepsilon$ or $C = \varepsilon$ with ε very small, one could approximate these special cases. Making this restriction allow us to easily clasify the types of transformation that can be allowed. With this in mind we assume that $A \neq 0 \neq C$ and, consequently, the transformations can be of two types: affine transformations such that $AC > 0$ (let us call them Type I) or affine transformations such that $AC < 0$ (let us call them Type II).

3.3.5.1 Affine transformations such that $AC > 0$ (Type I)

These are such that π_T intersects the plane $b = 0$ in the control space on a line $a = (D - Cc)/A$, and the gradient $-C/A$ is negative (see Figure 3.13).



(A) **(B)**

Figure 3.13: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $AC > 0$. (A) Intersection of the plane $\pi_T = \{(a, b, c) \in \mathbb{R}^3 : a + b + c = -1\}$ and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) In orange, the intersection $\mathcal{B}_1 \cap \{(a, 0, c) \in \mathbb{R}^3\}$. In purple, the intersection $\mathcal{B}_2 \cap \{(a, 0, c) \in \mathbb{R}^3\}$. In cyan, the intersection $\pi_T \cap \{(a, 0, c) \in \mathbb{R}^3\}$. As we can see, the line has negative gradient.*

Since a tristable region needs to be present in the intersection $\pi_T \cap \mathcal{B}$, this forces

$$D/A < 0. \tag{3.30}$$

On the other hand, because of the experimental results for the Notch null with 2×EGF mutant (See experiment (5) in Table 3.1), and $lin-12$ gain-of-function experiments in Sternberg & Horvitz (1989) (which can be considered as $NOTCH$

high), we constrain the system so that EGF high or $NOTCH$ high drive the system out of the tristable region. Therefore,

$$m_{31}, m_{32} > 0 \quad \text{and} \quad m_{11}, m_{12} < 0. \tag{3.31}$$

Examples of Type I transformations that satisfy the constraints in Equations 3.30 and 3.31 are given in Figure 3.14.

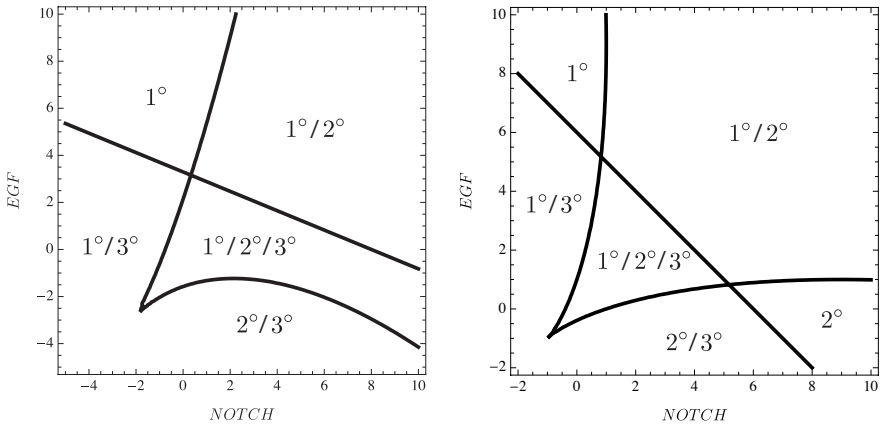


Figure 3.14: *Two examples of bifurcation sets in the signal space given by the intersections $T^{-1}(\pi_T \cap \mathcal{B})$ for two different Ts of Type I satisfying constraints in Equations 3.30 and 3.31. As in Figure 3.12, different regions are labelled with the corresponding attractors.*

Equations 3.26, 3.27, 3.28 and 3.31 allow us to rewrite three parameters of the transformation as functions of the other parameters, lowering the levels of freedom. In particular:

$$m_{31} = \sqrt{1 - m_{11}^2 - m_{21}^2} \tag{3.32}$$

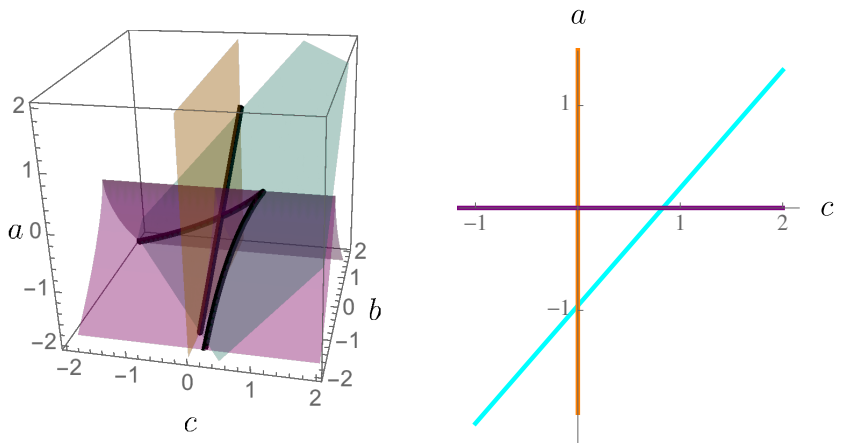
which let us write m_{22} and m_{32} as the solutions of the system

$$\begin{cases} m_{12}^2 + m_{22}^2 + m_{32}^2 = 1 \\ m_{11}m_{12} + m_{21}m_{22} + m_{31}m_{32} = 0 \end{cases} \tag{3.33}$$

3.3.5.2 Affine transformations such that $AC < 0$ (Type II)

These are such that π_T intersects the plane $b = 0$ on a line $a = (D - Cc)/A$, and the gradient $-C/A$ is positive (see Figure 3.15).

Since a tristable region needs to be present in the intersection $\pi_T \cap \mathcal{B}$, this forces



(A) **(B)**

Figure 3.15: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $AC < 0$. (A) Intersection of the plane $\pi_T = \{(a,b,c) \in \mathbb{R}^3 : a - b - c = -1\}$ and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) In orange, the intersection $\mathcal{B}_1 \cap \{(a,0,c) \in \mathbb{R}^3\}$. In purple, the intersection $\mathcal{B}_2 \cap \{(a,0,c) \in \mathbb{R}^3\}$. In cyan, the intersection $\pi_T cap\{(a,0,c) \in \mathbb{R}^3\}$. As we can see, the line has positive gradient.*

$$D/A < 0. \tag{3.34}$$

For the same reasons as before, we constrain the system so that EGF high or $NOTCH$ high drive the system out of the tristable region. Therefore,

$$m_{31}, m_{32} > 0 \quad \text{and} \quad m_{11}, m_{12} > 0. \tag{3.35}$$

Examples of Type II transformations that satisfy constraints in Equations 3.34 and 3.35 are given in Figures 3.16.

Equations 3.26, 3.27, 3.28 and3.31 again allow us to rewrite the parameters m_{31}, m_{22} and m_{32} of the transformation as functions of the other parameters, lowering the levels of freedom (See Equations 3.32 and 3.33).

Considering the system defined in Equation 3.19 and the linear transformations with the corresponding constraints, we can now present the proposed model that will describe the process of differentiation of the VPCs.

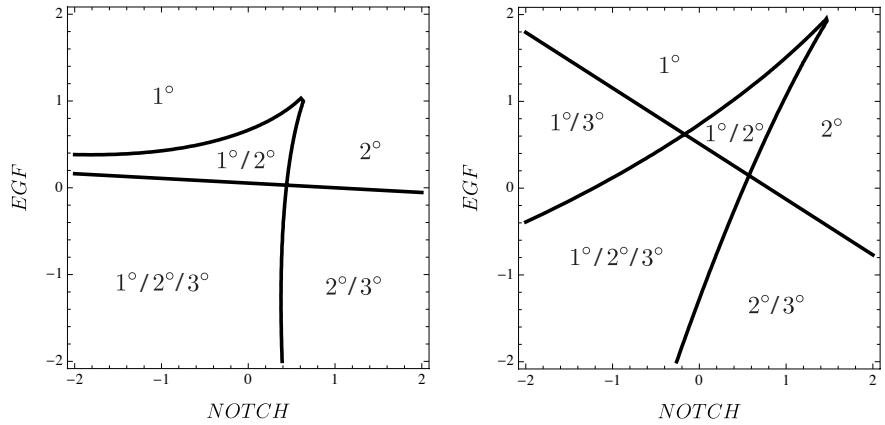


Figure 3.16: *Two examples of bifurcation sets in the signal space given by the intersections $T^{-1}(\pi_T \cap \mathcal{B})$ for two different Ts of Type II satisfiying constraints in Equations 3.34 and 3.35. As in Figure 3.12, different regions are labelled with the corresponding attractors.*

3.4 Model implementation and numerical simulations

As explained previously, we model the development of three VPCs (P4.p, P5.p and P6.p), since the pattern is always symmetrical around the AC. Our proposed model is described by the following set of equations, with (x_1, y_1) the coordinates representing the state of differentiation of P4.p, (x_2, y_2) the coordinates representing the state of P5.p and (x_3, y_3) the coordinates representing the state of P6.p:

$$\begin{cases} \frac{dx_1}{dt} &= \frac{1}{\tau}[\chi(y_1)(-4x_1^3 - 2a_1x_1 - b_1) - (1-\chi(y_1))x_1] + \eta_1(t) \\ \frac{dy_1}{dt} &= \frac{1}{\tau}[y_1((y_1 - M)^2 + c_1)] + \eta_2(t) \\ \frac{dx_2}{dt} &= \frac{1}{\tau}[\chi(y_2)(-4x_2^3 - 2a_2x_2 - b_2) - (1-\chi(y_2))x_2] + \eta_3(t) \\ \frac{dy_2}{dt} &= \frac{1}{\tau}[y_2((y_2 - M)^2 + c_2)] + \eta_4(t) \\ \frac{dx_3}{dt} &= \frac{1}{\tau}[\chi(y_3)(-4x_3^3 - 2a_3x_3 - b_3) - (1-\chi(y_3))x_3] + \eta_5(t) \\ \frac{dy_3}{dt} &= \frac{1}{\tau}[y_3((y_3 - M)^2 + c_3)] + \eta_6(t) \end{cases} \tag{3.36}$$

where, in order to work with smooth functions, we changed $\mathbb{1}_{\mathbb{R}_{\leq 0}}$ to $\chi(y)$, a sigmoidal function of the form:

$$\chi(y) = \frac{1 + \tanh\left(H\left(-y + rM\right)\right)}{2}, \tag{3.37}$$

where we choose $r = \frac{\sqrt{2}-1}{2\sqrt{2}}$.

The values of the control parameters for each cell are given by:

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = T(EGF, NOTCH) = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{pmatrix} \begin{pmatrix} EGF_i \\ NOTCH_i \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \tag{3.38}$$

where, as explained in Subsection 3.3.5, T is either of Type I or II, satifying the constraints in Table 3.2, where the definitions of A, C and D and the constraints are explained in Subsection 3.3.5.

Table 3.2: *Table of mathematical constraints on the affine transformation T of Type I and II from the signal space to the control space.*

Constraint	**For T of Type**
$A \neq 0$	I or II
$C \neq 0$	I or II
$8q_1^3 + 27q_2^2 < 0$	I or II
$q_3 < 0$	I or II
$AD < 0$	I or II
$m_{31} = \sqrt{1 - m_{11}^2 - m_{21}^2}$	I or II
m_{22}, m_{32} solutions of $\begin{cases} m_{12}^2 + m_{22}^2 + m_{32}^2 = 1 \\ m_{11}m_{12} + m_{21}m_{22} + m_{31}m_{32} = 0 \end{cases}$	I or II
$m_{31}, m_{32} > 0$	I or II
$AC > 0$	I
$m_{11}, m_{12} < 0$	I
$AC < 0$	II
$m_{11}, m_{12} > 0$	II

Following the approach by (Corson & Siggia 2012, 2017), we write:

$$EGF_1 = s\gamma^2, \quad EGF_2 = s\gamma, \quad EGF_3 = s, \tag{3.39}$$

which derivation follows from the diffusion of EGF morphogen ($\gamma < 1$), (see Supplemental material of Corson & Siggia (2012)). The parameter s is a scaling factor.

Inspired by Corson & Siggia (2012, 2017), we define the $NOTCH$ parameter to be proportional to the sum of the autocrine Notch signal of the cell (signal pro-

duced by the cell itself) and the paracrine Notch signal received by the cell (produced by the neighboring cells). The autocrine signal is multiplied by a parameter α which parametrises the relative importance of autocrine and paracrine signalling. As in Corson & Siggia (2017), the $NOTCH$ level also takes into account the downregulation of Notch signal in 1° fated cells, as suggested in Shaye & Greenwald (2002). This downregulation is related to the cell's production of Notch signalling, scaled by a parameter l_d which defines the strength of such downregulation. Therefore, we define the $NOTCH$ levels as

$$NOTCH_1 = l(1 - l_d L(x_1, y_1))(\alpha L(x_1, y_1) + L(x_2, y_2)), \tag{3.40}$$

$$NOTCH_2 = l(1 - l_d L(x_2, y_2))(L(x_1, y_1) + \alpha L(x_2, y_2) + L(x_3, y_3)), \tag{3.41}$$

$$NOTCH_3 = l(1 - l_d L(x_2, y_2))(\alpha L(x_3, y_3) + 2L(x_2, y_2)), \tag{3.42}$$

where the parameter l is a scaling factor. The function L represents the Notch signal emitted by a cell, which depends on the current state of the cell. L should be increasing as the state of the cell approaches the attractor corresponding to 1° fate; therefore if (x_1^*, y_1^*) are the coordinates of the attractor corresponding to 1° fate, we write:

$$L(x, y) = L_1(x)L_2(y) = \frac{1 + \tanh(H_1(M_1 - x))}{2} \frac{1 + \tanh(H_2(M_2 - y))}{2}, \tag{3.43}$$

where we could fix $M_1 = x_1^*$ and $M_2 = y_1^*$. However, we decide to allow the parameters M_1 and M_2 to be fitted.

Regarding the noise, we consider η_i to be a white noise with variance $2\sigma_{dif}$, to account for intrinsic variability:

$$\langle \eta_i(t)\eta_j(t') \rangle = 2\sigma_{dif}\delta_{ij}\delta(t - t'). \tag{3.44}$$

The development of the VPCs is then defined by a 6-dimensional system of stochastic differential equations with the eighteen parameters in Table 3.3.

After fixing certain experimental conditions represented by the value of a parameter vector, the system in (3.36) allows us to simulate the experimental outcome following the steps given in the following subsections of this section.

Table 3.3: *Table of parameters of the proposed model for vulval development.*

Parameters		
$1/\tau$	q_1	M_2
H	q_2	s
M	q_3	l
m_{11}	H_1	α
m_{12}	M_1	l_d
m_{21}	H_2	σ_{dif}

3.4.1 Initial condition

In order to perform a simulation, first, it is necessary to choose an initial condition that will represent the state of the VPCs at the start of the experiment. We assume that the VPCs are initially equivalent and that their initial condition should not depend on the signals they will receive. Since in the absence of signals they take the 3° fate, we assume that their initial condition should lie in the basin of attraction of the attractor corresponding to the 3° fate.

It is also reasonable to incorporate some variability in the initial state, so we decide to choose as initial condition of the full system, the stationary distribution around the equilibrium of the following system of SDEs

$$\begin{cases} \frac{dx}{dt} &= \frac{1}{\tau}[\chi(y)(-4x^3 - 2a_0x - b_0) - (1-\chi(y))x] + \eta_{01}(t) \\ \frac{dy}{dt} &= \frac{1}{\tau}[y((y-M)^2 + c_0)] + \eta_{02}(t) \end{cases} \tag{3.45}$$

where $a_0 = q_1, b_0 = q_2, c_0 = q_3$ are obtained from equation (3.38) by setting $EGF = 0$ and $NOTCH = 0$ and η_{01} and η_{02} represent independent white noises with variance $2\sigma_{dif}$.

As initial condition of the system in Equation 3.45 we choose a normal distribution centered at $(x(0), y(0)) = (0, M)$ with zero covariance matrix, to make sure that the initial condition stays in the basin of attraction of the tertiary fate.

We decide to approximate such stationary distribution numerically. In order to do that, we considered two approaches. As a first attempt, we approximated such distribution by running N simulations ($N = 100, 1000$) for a long enough period of time by taking advantage of the Euler-Maruyama method (See Appendix (A.1)). The stationary distribution was approximated by the distribution of the end points of those N simulations. As a seccond approach, we approximated the solution of the

SDE based on a similar method to derive the Linear Noise Approximation (LNA) (see Appendix A.3), and computed the solution for a long enough period of time. In order to get the same accuracy with the two methods, the first approach took twice as much time as the second approach, so we decided to use the second one.

3.4.2 Simulation procedure

Equation 3.36 is solved by using the Euler-Maruyama method (See Appendix (A.1)). We draw N initial conditions from the initial distribution computed as we explained before and simulate N random walks with the Euler Maruyama method. The system in Equation 3.36 is simulated from $t = 0$ to $t = 1$, with time step $dt = 0.02$, with signals active, to account for the competence period in which the VPCs respond to signals. Then, the system is continued from $t = 1$ to $t = 3$, with $dt = 0.02$, with signals off to account for the post-competence period. We consider this post-competence period because, as explained in Subsection 3.3.5, in order for cells to be specified, they should keep their fates even if they do not receive any signals. During this post-competence period we consider an exponential decay of the signals (we found anomalous solutions if we instantaneously set $EGF = NOTCH = 0$, and it is also more biologically reasonable). Therefore, during the post-competence period, we replace s and l in Equations 3.39, 3.40, 3.41 and 3.42 by:

$$s^{PC}(t) = se^{-\lambda_E t}, \tag{3.46}$$

$$l^{PC}(t) = le^{-\lambda_N t}. \tag{3.47}$$

The parameters λ_E and λ_N are fixed to be equal to 3 so that, at $t = 3$, $s^{PC}(3), l^{PC}(3) \approx 0$.

At time $t = 3$, for each one of the N simulations, the fates of the three VPCs are scored as will be explained in Subsection 3.4.3. The simulated outcome is given by the proportion of times that each cell took each fate. Therefore, the simulated outcome can be sumarised in a matrix

$$\mathcal{D}^{sim} = \begin{pmatrix} p_{11}^{sim} & p_{12}^{sim} & p_{12}^{sim} & p_{14}^{sim} \\ p_{21}^{sim} & p_{22}^{sim} & p_{22}^{sim} & p_{24}^{sim} \\ p_{31}^{sim} & p_{32}^{sim} & p_{32}^{sim} & p_{34}^{sim} \end{pmatrix} \tag{3.48}$$

where for $j = 1, 2, 3$ the value p_{ij}^{sim} is the proportion of times, in the N simulations, that P$(i + 3)$.p took fate j°. For $j = 4$ the value p_{ij}^{sim} is the proportion of times, in the 64 simulations, that P$(i + 3)$.p could not be assigned a fate (See Subsection

3.4.3 for details). We checked the sensitivity of $\mathcal{D}^{sim}$ with respect to the number of simulations and we found that $N = 64$ gave a reasonable approximation compared to $N = 100$ or $N = 200$, therefore we chose $N = 64$.

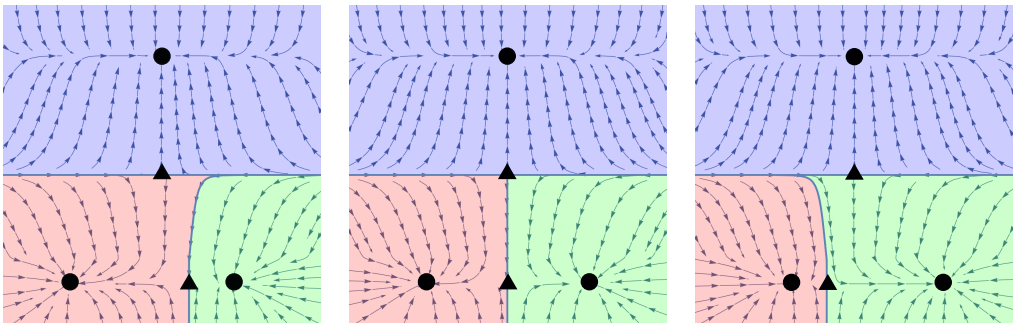
3.4.3 Fate assignment

Given one simulation of the system, the goal is to compute the fate of each cell. As explained in Subsection 3.4.2, we score the fates at time $t = 3$. At that time, the signals are switched off for the three VPCs, so they *see* the same fixed landscape defined by the system of equations:

$$\begin{cases} \frac{dx}{dt} &= \frac{1}{\tau}[\chi(y)(-4x^3 - 2a_{PC}x - b_{PC}) - (1 - \chi(y))x] \\ \frac{dy}{dt} &= \frac{1}{\tau}[y((y - M)^2 + c_{PC})] \end{cases} \tag{3.49}$$

where $\chi(y)$ is defined in Equation 3.37, $a_{PC} = q_1$, $b_{PC} = q_2$, $c_{PC} = q_3$, such that it contains the three attractors (one for each fate) (See explanation after Equation 3.29 in Subsection 3.3.5).

Therefore, depending on the value of $b_{PC} = q_2$, the basins of attraction in the landscape will look like the ones in Figure 3.17.



(A) $q_2 > 0$ **(B)** $q_2 = 0$ **(C)** $q_2 < 0$

Figure 3.17: *Possible landscapes for post-competence period for different values of q_2, when the control parameters take values $a_{PC} = q_1$, $b_{PC} = q_2$, $c_{PC} = q_3$. Arrows represent the flow defined by the system in Equation 3.49. Basins of attraction are coloured blue, green and red from attractors corresponding to the 3°, 2° and 1° fates, respectively.*

In order to assign a fate to a VPC in one simulation, we find the basin of attraction in which the trajectory lays at time $t = 3$, and allocate the corresponding fate.

Let us call (x_i, y_i) coordinates of a cell at time $t = 3$ for a particular simulation. In order to find the basin of attraction in which it lies we do the following:

1. Suppose $y_i > M - \sqrt{-c}$. As we see in Figure 3.18, the basin of attraction

corresponding to the 3° fate is the region of the state space defined by $\{(x, y) \in \mathbb{R}^2 : y > M - \sqrt{-c}\}$, where $M - \sqrt{-c}$ is the y-coordinate of the yellow saddle in Figure 3.18.

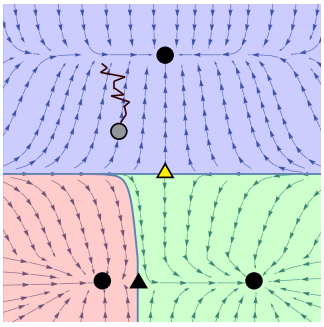


Figure 3.18: *Example of an assignment of the fate 3° to a VPC in a simulation. Arrows represent the flow defined by the system in Equation 3.49. Basins of attraction are coloured blue, green and red from attractors corresponding to the 3°, 2° and 1° fates, respectively. The end point (gray circle) of a simulated trajectory of a VPC during the post-competence period (black path) lays in the basin of attraction of 3° fate (blue region).*

Therefore, if $y_i > M - \sqrt{-c}$, then we assign the 3° fate to this VPC in such a simulation.

2. If $y_i = M - \sqrt{-c}$, then the point lies on the boundary between two basins of attraction. We consider that we cannot assign a fate to the VPC in this case.

3. If $y_i < M - \sqrt{-c}$ then the point lies in the basin of attraction of either 1° or 2° fate.

   (a) If $q_2 = 0$, the saddle between the red and green regions in Figure 3.17 has coordinates $(0, 0)$ and it is half way between the attractors corresponding to fates 1° and 2°. In this case, the basin of attraction corresponding to 1° fate is defined by the set of points with $x < 0, y < M - \sqrt{-c}$. Moreover, the basin of attraction corresponding to the 2° fate is defined by the set of points given by $x > 0, y < M - \sqrt{-c}$. Consequently, we assign fate 1° (resp. 2°) to such a VPC if $x_i < 0$ (resp. $x_i > 0$).

   (b) If $q_2 \neq 0$, we observe that the stable manifold of the saddle between the red and green regions, which serves as boundary between the two basins of attraction, bends towards the closer of the two attractors (See Figure 3.17).

   Consider the case $q_2 < 0$ (similarly if $q_2 > 0$). In this case, the landscape would look like the leftmost landscape in Figure 3.17. Let $(x_1^*, 0)$ be the

coordinates of the attractor corresponding to the 1° fate, $(x_2^*, 0)$ be the coordinates of the attractor corresponding to the 2° fate, $(x_s^*, 0)$ be the coordinates of the saddle determining the boundary between the basins of attraction of 1° and 2° fates, and $(0, M - \sqrt{-c_{PC}})$ be the coordinates of the saddle determining the basin of attraction of the 3° fate. We assign the fate as follows:

i. If $(x_i, y_i) \in C_2 = \{(x, y) \in \mathbb{R}^2 : \|(x, y) - (x_2^*, 0)\| \leq R\|(x_s^*, 0) - (x_2^*, 0)\|$, i.e. (x_i, y_i) is in the dark green circle arround the attractor corresponding to 2° fate in Figure 3.19), we assign the fate 2°.

ii. If $(x_i, y_i) \in C_1 = \{(x, y) \in \mathbb{R}^2 : \|(x, y) - (x_1^*, 0)\| \leq Smin\{\|(x_s^*, 0) - (x_1^*, 0)\|, M - \sqrt{-c_{PC}}\}$ i.e. (x_i, y_i) is in the dark red circle arround the attractor corresponding to 1° fate in Figure 3.19), we assign fate 1°.

iii. Otherwise, we solve the system of ordinary differential equations in Equation 3.49 until the trajectory crosses either C_1 or C_2, assigning the corresponding fate.

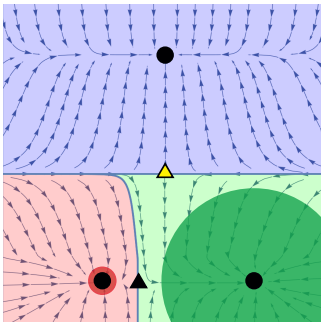


Figure 3.19: *Assignment of* 1° *or* 2° *fates when* $q_2 < 0$. *Arrows represent the flow defined by the system in Equation 3.49. Basins of attraction are coloured blue, green and red from attractors corresponding to* 3°, 2° *and* 1° *fates, respectively. Dark green region corresponds to the set* $C_2 = \{(x, y) \in \mathbb{R}^2 : \|(x, y) - (x_2^*, 0)\| \leq R\|(x_s^*, 0) - (x_2^*, 0)\|$. *Red green region corresponds to the set* $C_1 = \{(x, y) \in \mathbb{R}^2 : \|(x, y) - (x_1^*, 0)\| \leq Smin\{\|(x_s^*, 0) - (x_1^*, 0)\|, M - \sqrt{-c_{PC}}\}$

Where we take $R = 0.9$ and $S = 10^{-3}$.

3.4.4 Simulation of mutants

If we call

$$\Theta_{WT} = (1/\tau, H, M, m_{11}, m_{12}, m_{21}, q_1, q_2, q_3, H_1, M_1, H_2, M_2, s, l, \alpha, l_d, \sigma_{dif}),$$

the parameter vector that represents WT conditions in the model, we can simulate all the mutants in Table 3.1 by changing the corresponding parameters the following way.

3.4.4.1 *let-23* mosaic

In this mutant worm, there are no EGF receptors in P5.p. Therefore we simulate this mutant with the same parameter vector Θ_{WT} as the WT but we change the parameter that controls the EGF signal for P5.p. In particular, we set $s = 0$ only in P5.p, to get the parameter vector $\Theta_{(2)}$.

3.4.4.2 Half dose of *lin-3*

In this case, the EGF ligand is reduced by half with respect to the WT case in all the VPCs. We can translate this into our model by taking the parameter $s/2$ instead of s in the WT parameter vector, giving the parameter vector $\Theta_{(3)}$.

3.4.4.3 Half dose of *lin-12*

In this case, the Notch receptor is reduced by half with respect to the WT case in all the VPCs. To do this we reduce half l in the WT parameter vector to get a parameter vector $\Theta_{(4)}$.

3.4.4.4 Notch null, 2× WT EGF

In this mutant, the VPCs lack the Notch receptor, and the EGF that they receive is twice as much the EGF signal that they receive in the WT worm. Taking this into account, we simulate this mutant by doubling s and setting l to zero in the WT parameter vector to get $\Theta_{(5)}$.

3.4.4.5 No notch signalling, WT EGF

As in the previous case, the VPCs do not receive any Notch signal because they lack the Notch receptor. However, in this case, EGF signal is the same as in the WT

case. We can translate this into our model by setting l to zero in the WT parameter vector to obtain $\Theta_{(6)}$.

3.4.4.6 EGF overexpression

In this case, the VPCs receive a higher EGF signal but the fold change is unknown. We add a parameter $s_E > 0$ to the model, that will represent the fold change of EGF mutant in this mutant with respect to the WT case. Therefore, to get the parameter vector $\Theta_{(7)}$ with which we simulate this experiment, we multiply s by s_E in the WT parameter vector.

3.4.5 AC ablation mutants

These mutants correspond to experiments in which the AC is ablated (i.e. EGF signal is removed) at different stages of the development of the worm. If we simulate the competence time of the WT case from $t = 0$ to $t = 1$, we simulate the competence time of these mutants by taking the parameter vectors:

$$\begin{cases} \Theta_{WT} & t \in [0, t_{ACAb}] \\ \Theta_{ACAb} & t \in (t_{ACAb}, 1] \end{cases}$$

where Θ_{ACab} is obtained by setting $s = 0$ in the WT parameter vector, meaning that EGF is removed, and t_{ACAb} represents the developmental stage. Since we are not able to find the numerical timing of each developmental stage in the literature we follow the approach taken by Corson & Siggia (2012) where t_{ACAb} takes the values in the following table

Developmental Stage	t_{ACAb}
Let. L2	0.02
eL3	0.32
DU	0.44
VU	0.56
3°	0.68
2-cell	0.8

(3.50)

Given a set of parameters, we can now compute the proportion of cells that take the different fates and we can compare the simulated outcome to the experimental outcome. The next step is to find the parameters that allow us to reproduce experimental data.

Chapter 4

Parameter estimation for vulval development model

4.1 Introduction

In the previous chapter we have introduced a model for the process of vulval development in *C. elegans*. Given a set of parameters, we can simulate different experiments and predict the proportion of cells that would adopt each fate. Our aim is to find the set of parameters that allows the model to reproduce the experimental data given by Corson & Siggia (2012) (Table 4.1). This table contains the probabilities of each VPC (P4.p, P5.p, P6.p) acquiring each fate (1°, 2°, 3°) under thirteen experimental conditions. For experiments $8-13$ the probabilities for only two VPCs are observed. In total, Table 4.1 consists of $3\times3\times7+2\times3\times6=99$ data points. We will denote by $p_{e,f,c}$ the experimental probability of cell c taking fate f under experimental condition e, the values of which are given in Table 4.1. Similarly, $p^{sim}_{e,f,c}(\boldsymbol{\theta})$ is the simulated probability of cell c taking fate f under experimental condition e given parameters $\boldsymbol{\theta}$. Our goal is to estimate the parameters $\boldsymbol{\theta}^*$ such that the 99-dimensional vector of simulated data $\boldsymbol{X}(\boldsymbol{\theta})=(p^{sim}_{e,f,c}(\boldsymbol{\theta}^*))$ reproduces the vector of real data $\boldsymbol{X}_0=(p_{e,f,c})$.

We take a Bayesian approach to this problem. With this approach, as explained in Section 2.4 of Chapter 2, parameters are treated as **random variables** and hence they have a probability distribution. If we denote the parameter vector by $\boldsymbol{\theta}$, the posterior distribution satisfies:

$$\pi(\boldsymbol{\theta}\mid\boldsymbol{X}_0)\propto\pi(\boldsymbol{\theta})\mathcal{L}(\boldsymbol{X}_0\mid\boldsymbol{\theta}). \tag{4.1}$$

However, it is not possible to find an analytical expression for the likelihood function of the stochastic model proposed in Chapter 3. Therefore, it is necessary to use computational methods, such as those discussed in Section 2.4, to approximate the posterior distribution of interest.

One possible approach is to approximate the likelihood function, $\mathcal{L}(\boldsymbol{X}_0\mid\boldsymbol{\theta})$, as in Corson & Siggia (2012, 2017). In a similar way to regression or least-square

Table 4.1: *Table of experimental data obtained from the literature (Corson & Siggia 2012), fitted with the model. Fates for P4.p and P8.p (or P5.p and P7.p) have been averaged, since we assume that the pattern is symmetrical around the anchor cell. Developmental stages: (Let. L2) lethargic L2; (eL3) early L3; (DU) Dorsal Uterine precursor cells dividing or divided once; (VU) Ventral Uterine precursor cells dividing or divided once; (3°) 3° cells have divided; (2-cell) all Pn.p cells have divided once.*

	Experiment	VPC fates (% 1°, 2° 3°)		
		P4.p	P5.p	P6.p
	Wild-type outcomes under reduced signalling			
(1)	Wild type	0, 0, 100	0, 100, 0	100, 0, 0
(2)	*let-23* mosaic		wild type	
	(no EGF receptors in P5/7.p)			
(3)	Half dose of *lin-3*		wild type	
	(Half EGF ligand)			
(4)	Half dose of *lin-12*		wild type	
	(Half Notch receptor)			
	Phenotypes under reduced Notch/excess EGF			
(5)	Notch null, and 2×WT EGF (2 ACs)	0, 0, 100	100, 0, 0	100, 0, 0
	(No Notch receptor, twice as much WT EGF ligand)			
(6)	No Notch signalling, WT EGF	0, 0, 100	0, 0, 100	100, 0, 0
	(No Notch receptor, WT EGF ligand)			
(7)	EGF overexpression	18,46,36	45.5,54.5,0	96,4,0
	(overexpression of EGF ligand)			
	Phenotypes following anchor cell ablation			
(8)	L2 lethargus	-	0, 0, 100	0, 0, 100
(9)	Early L3	-	1.5, 21, 77.5	18, 18, 64
(10)	DU divided	-	0, 54.63, 45.37	31, 38, 31
(11)	VU divided	-	4, 90, 6	52, 48, 0
(12)	3° divided	-	1, 99, 0	65, 35, 0
(13)	2-cell stage	-	1, 99, 0	93, 7, 0

procedures, they define the likelihood through the error between the predicted data $\boldsymbol{X}(\boldsymbol{\theta}) = (p^{sim}_{e,f,c}(\boldsymbol{\theta}))$ and the experimental data $\boldsymbol{X}_0 = (p_{e,f,c})$. This likelihood function takes the form:

$$\mathcal{L}(\boldsymbol{X}_0 \mid \boldsymbol{\theta}) = e^{-\frac{\chi(\boldsymbol{\theta},\boldsymbol{X}_0)^2}{2}} \tag{4.2}$$

where in Corson & Siggia (2012)

$$\chi(\boldsymbol{\theta}, \boldsymbol{X}_0)^2 = N \sum_{e,f,c} \left(p_{e,f,c} - p^{sim}_{e,f,c}(\boldsymbol{\theta})\right)^2 \tag{4.3}$$

and in Corson & Siggia (2017)

$$\chi(\boldsymbol{\theta}, \boldsymbol{X}_0)^2 = \frac{N^2}{4} \sum_{e,f,c} \frac{\left(p_{e,f,c} - p^{sim}_{e,f,c}(\boldsymbol{\theta})\right)^2}{N p^{sim}_{e,f,c}(\boldsymbol{\theta}) + \frac{1}{2}}. \tag{4.4}$$

One can see that the likelihood is maximised when $\boldsymbol{\theta}$ reproduces the experimental data well.

This approximation of the likelihood facilitates the use of MCMC methods, specifically the adaptive Metropolis (AM) algorithm explained in Section 2.4 of Chapter 2.

The application of MCMC methods requires the parameter vector to be initialised and the authors use an approximation of $\boldsymbol{\theta}^* = \arg\max_{\boldsymbol{\theta}}\{\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0)\}$ to do this. The definition of the likelihood function in Equations (4.2)–(4.4) and a combination of flat, normal and exponential priors, allowed them to write the logarithm of the posterior as a sum of squares, which is maximised using the **Levenberg-Marquardt method** (Vetterling et al. 1989). This iterative algorithm returns an approximation of the global maximum $\boldsymbol{\theta}^*$, provided that the initial condition is somewhat close to it. Since this method needs the function $\log(p(\boldsymbol{\theta}|X_0))$ to be deterministic and continuous, the noise realisations are fixed. The authors claim that a global maximum is reached by starting the algorithm with different initial parameter values. Once the initial conditions are reached, the authors use AM algorithm to explore the vicinity of the global optimum.

This approach has some questionable assumptions. First of all, it is difficult to justify that the optimum parameter $\boldsymbol{\theta}^*$ found with the Levenberg-Marquardt method is the global optimum. Since the noise has to be fixed, this optimum will depend on the noise realisations. It also depends on the initial condition taken and the number of initial conditions that would need to be considered is highly dependent on the number of fitted parameters. Moreover, even if the algorithm is started from

many different initial values and all the runs converge to the same result, that does not guarantee that the global optimum has been found. This has a large impact on the estimated posterior because their implementation of MCMC only searches in the vicinity of the initial condition.

A second approach is to take advantage of likelihood-free techniques to approximate the posterior distribution π. In particular, as we introduced in Section 2.4 of Chapter 2, **Approximate Bayesian computation** (ABC) methods have been succesfully applied to a wide range of intractable likelihood problems over the past eighteen years (Marin et al. 2012). ABC methods give more flexibility for the comparison between the experimental and simulated data, which is important in the absence of an analytical expression for the likelihood. Moreover, the ABC SMC algorithm has computational advantages since it can be parallelised, as we will explain in Subsection 4.2.6. This is especially important when the parameter space that needs to be explored by the fitting algorithm is high dimensional.

For these reasons we decided to use the likelihood-free approach. In this chapter, we will provide an explanation of the implementation of the ABC method for our purposes and will present some results for the parameter fitting of the model proposed in Chapter 3.

4.2 ABC SMC Implementation

We take advantage of sequential Monte Carlo ABC (ABC SMC) algorithm introduced in Section 2.4 of Chapter 2. This algorithm samples sequentially from a sequence of intermediate distributions that increasingly resemble the posterior distribution given in equation 4.1.

For this implementation, we first need to study the dependence on the parameters of our model. We will study the range of the parameters, consider some constraints that the model and certain mutants in Table 4.1 impose on the parameters, and provide the corresponding priors.

Secondly, we need to define a distance function d that will measure the *similarity* between the experimental data set and the corresponding simulated data set, i.e. the **goodness of fit**.

Defining this distance allows us to perform a sensitivity analysis to study which parameters affect the outcome the most. This will enable us to reduce the dimensionality of the space that the fitting algorithm will explore as we can fix the parameters that the model is least sensitive to.

For the implementation of the ABC SMC method (as explained in Section

2.4 of Chapter 2), we also need to define a decreasing sequence of thresholds $\{\varepsilon_t\}_{t\geq 0}$ that will determine the maximum distance between the experimental data and the simulated data in each step t of the algorithm. These in turn define the intermediate distribution p_{ε_t} from which the algorithm samples at step t. Remember that these intermediate distributions have the form

$$\{\pi_t\}_{0\leq t\leq T} = \{\pi(\boldsymbol{\theta} \mid d(\boldsymbol{X}(\boldsymbol{\theta})), \boldsymbol{X}_0) \leq \varepsilon_t)\}_{0\leq t\leq T} \tag{4.5}$$

Also, we need to choose the number N of particles sampled from the distribution p_{ε_t} at each step t.

Finally, we need to provide the perturbation kernels $\{K_t(\cdot \mid \cdot)\}_{t=1}^{N}$ that will set the limits of exploration of the parameter space in each step.

We will discuss these points further in the following subsections.

4.2.1 Range

The range of values that each parameter can take is given on Table 4.2. These ranges are imposed by the model definition and constraints in Table 3.2.

Table 4.2: *Tables of ranges of the parameters in Type I and Type II models.*

Parameters	Range	
	Type I	**Type II**
$1/\tau$	$(0,\infty)$	$(0,\infty)$
H	$(0,\infty)$	$(0,\infty)$
M	$(0,\infty)$	$(0,\infty)$
m_{11}, m_{12}	$[-1,0]$	$[0,1]$
m_{21}	$[-1,1]$	$[-1,1]$
q_1, q_2	$\mathbb{R}$	$\mathbb{R}$
q_3	$(-\infty,0)$	$(-\infty,0)$
γ	$[0,1]$	$[0,1]$

Parameters	Range	
	Type I	**Type II**
H_1, H_2	$(0,\infty)$	$(0,\infty)$
M_1, M_2	$\mathbb{R}$	$\mathbb{R}$
s, l	$(0,\infty)$	$(0,\infty)$
α	$(0,\infty)$	$(0,\infty)$
l_d	$[0,1]$	$[0,1]$
s_E	$(0,\infty)$	$(0,\infty)$
σ_{dif}	$(0,\infty)$	$(0,\infty)$

4.2.2 Constraints on the parameters

In Table 3.2, we sumarised the constraints that the transformation from the signal space to the control space imposes on the parameters. In the following subsections

we will outline three other constraints imposed on the parameters by the model and the data in Table 4.1.

4.2.2.1 Constraint imposed by the sigmoidal function $\chi(y)$

As introduced in Section 3.4 of Chapter 3, we approximate the indicator function $\mathbb{1}_{\mathbb{R}_{\leq 0}}(y)$ by a sigmoidal function that has the form

$$\chi(y) = \frac{1 + \tanh\left(H\left(-y + rM\right)\right)}{2} \tag{4.6}$$

where we choose $r = \frac{\sqrt{2}-1}{2\sqrt{2}}$.

Focusing on the deterministic system at first, the trajectory corresponding to the state of a VPC is moving on a landscape characterised by the equilibria of the system:

$$\begin{cases} \frac{dx}{dt} &= \frac{1}{\tau}[\chi(y)(-4x^3 - 2ax - b) - (1 - \chi(y))x] \\ \frac{dy}{dt} &= \frac{1}{\tau}[y((y - M)^2 + c)]. \end{cases} \tag{4.7}$$

This is similar to Equation 3.5 but we have substituted the indicator function by the sigmoidal function. The control parameters a, b, c are functions of the signals as we explained in Chapter 3.

As studied in Subsection 3.3.2, the critical points of the system in Equation 4.7 will have y-coordinates equal to 0, $M - \sqrt{-c}$ and $M + \sqrt{-c}$ (these two latter ones only when $c < 0$).

The constraint that we introduce here is firstly due to the fact that $\chi(0)$ needs to be approximately equal to 1, so that the values of the x-coordinates of the critical points come only from the cusp equation. Secondly, $\chi(M - \sqrt{-c})$ and $\chi(M + \sqrt{-c})$ need to be approximately equal to zero so that there is only one possible value of the x-coordinate for these critical points and it must be equal to zero.

Consequently, we need to impose:

1. $\chi(0) \approx 1$

   This condition can be translated into the condition $1 \geq \chi(0) \geq 1 - \varepsilon$ where ε is small.

   Substituting in Equation 4.6 we get:

$$\chi(0) = \frac{1 + \tanh\left(H\left(rM\right)\right)}{2} \geq 1 - \varepsilon \Leftrightarrow M \geq \frac{\operatorname{atanh}(1 - 2\varepsilon)}{rH}$$

This constrains the smallest value that the parameter $M > 0$ can take. This minimum value of M will depend on how accurate we would like the function $\chi(y)$ to be or, in other words, how small ε is, as well as on the steepness H of the sigmoidal function. We choose $\varepsilon = 10^{-12}/2$, and leave $H > 0$ as a parameter to be estimated. Therefore

$$M \geq M_{min}(H) = \frac{\text{atanh}(1 - 10^{-12})}{\frac{\sqrt{2}-1}{2\sqrt{2}} H} \tag{4.8}$$

See Figure 4.1 for a graphical explanation.

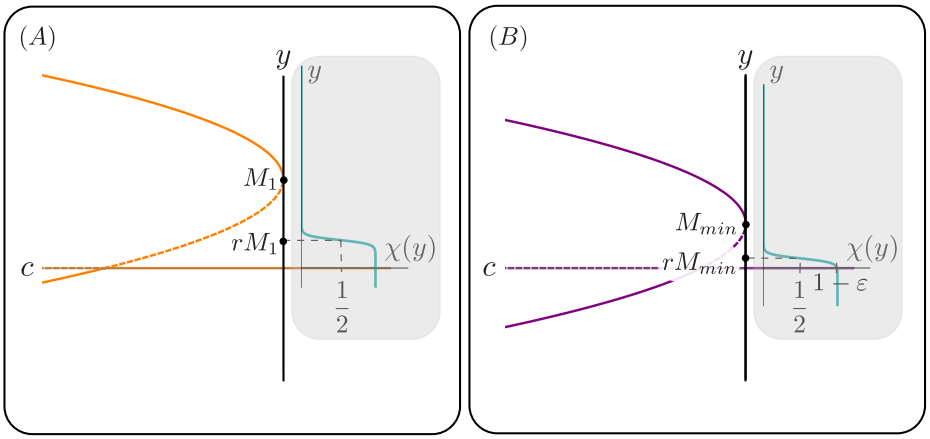


Figure 4.1: *Bifurcation diagram of the y-coordinate with respect to the control parameter c, together with a plot of $\chi(y)$, for $M = M_1$ (on the left plot (A)) and $M = M_{min}$ (on the right plot (B)). Fixed a value of H, the value of M "shifts" the graph of $\chi(y)$ along the y-axis, by changing the value of rM where $\chi(rM) = 1/2$. The minimum value M_{min} that M can take is such that $\chi(0) = 1 - \varepsilon$, otherwise $\chi(0)$ would be "too far" from 1.*

2. $\chi(y) \approx 0$ when $y > M - \sqrt{-c}$ for all M and c allowed.

   This constraint is necessary so that the value of $\chi(y)$ evaluated at the critical points $y = M - \sqrt{-c}$, $y = M + \sqrt{-c}$ is approximately zero for all values of M and c allowed.

   We know that the sigmoidal function is monotonically decreasing, therefore $\chi(M + \sqrt{-c}) \leq \chi(M - \sqrt{-c})$. This means that we only need to impose $\chi(M - \sqrt{-c}) \leq \varepsilon$, where ε takes the same value as before ($\varepsilon = 10^{-12}/2$).

   To achieve that, we restrict the value of c to be bigger than a value c_{min}, where c_{min} is such that $\chi(M - \sqrt{-c_{min}}) = \varepsilon$. If $c < c_{min}$ then $\chi(M - \sqrt{-c}) > \varepsilon$, because $\chi(y)$ is monotonically decreasing. Therefore we should avoid values $c < c_{min}$ (see Figure 4.2).

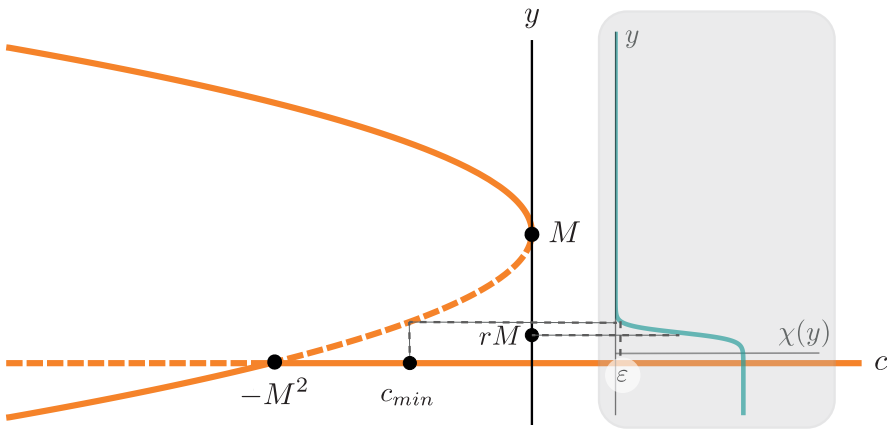


Figure 4.2: *Bifurcation diagram of the y coordinate together with a plot of $\chi(y)$. $\chi(y) = 1/2$ when $y = rM$. The value c_{min} is such that $\chi(M - \sqrt{c_{min}}) = \varepsilon$.*

We should therefore impose:

$$c > c_{min} = \min_c \{\chi(M - \sqrt{-c}) \leq \varepsilon\} \tag{4.9}$$

Now,

$$\chi(M - \sqrt{-c_{min}}) = \frac{1 + \tanh\left(H\left(-M + \sqrt{-c_{min}} + rM\right)\right)}{2} = \varepsilon$$

implies that

$$c > c_{min} = -\left(\frac{\text{atanh}(1 - 2\varepsilon)}{H} + rM - M\right)^2. \tag{4.10}$$

But in our model, the control parameter c is written as a function of the signals as described in Section 3.4 of Chapter 3. In particular:

$$c(\text{EGF,NOTCH}) = m_{31}EGF + m_{32}NOTCH + q_3.$$

Considering the forms of the EGF and $NOTCH$ functions (see Equations 3.39, 3.40, 3.41, 3.42, 3.43) and the fact that $m_{31}, m_{32} > 0$ (see Equation 3.2) we know that $c(\text{EGF,NOTCH}) \geq q_3$.

This leads us to the final condition:

$$q_3 > -\left(\frac{\text{atanh}(1 - 2\varepsilon)}{H} + rM - M\right)^2, \tag{4.11}$$

where $r = (\sqrt{2}-1)/2\sqrt{2}$ and $\varepsilon = 10^{-12}/2$.

4.2.2.2 Constraint imposed by mutant (5) Notch null, 2×WT EGF

For this mutant, as explained in Section 3.4, there is no Notch signal in any of the VPCs since they lack the Notch receptor. This means that, in our model, the $NOTCH$ coordinate is zero for all the cells at all times, i.e. $l = 0$. Since the EGF signal is always constant for all the cells throughout the experiment, the EGF values are also constants for all the cells ($EGF_3 = 2s$ for P6.p, $EGF_2 = 2s\gamma$ for P5.p and $EGF_1 = 2s\gamma^2$ for P4.p as explained in Section 3.4 of Chapter 3). Therefore, the control parameters for the cells in this experiment are fixed in time and are

$$\begin{pmatrix} a_i \\ b_i \\ c_i \end{pmatrix} = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{pmatrix} \begin{pmatrix} EGF_i \\ 0 \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \tag{4.12}$$

meaning that the landscape is constant in time for all the VPCs since the control parameters are fixed in time.

In Table 4.1 we can see that P5.p and P6.p adopt fate 1°. However, in our model, P5.p and P6.p could only adopt fate 1° if their trajectory escapes the basin of attraction of the attractor representing fate 3° when the signals are not present. This will only happen if the attractor corresponding to 3° fate bifurcates away for the values of the control parameters in Equation 4.12.

This gives us the equations

$$\begin{aligned} c_3 &= 2s + q_3 \geq 0 \\ c_2 &= 2s\gamma + q_3 \geq 0. \end{aligned} \tag{4.13}$$

However, since $\gamma \in (0,1)$ and $s > 0$, we only need to impose $2s\gamma + q_3 \geq 0$.

4.2.2.3 Constraint imposed by mutant (6) Notch null, WT EGF

As before, there is no Notch signal in any of the cells and EGF is constant and equal to the WT signal. This means that, in our model, the $NOTCH$ coordinate is zero for all the cells at all times, i.e. $l = 0$, and so it is EGF coordinate, which is equal to $EGF_3 = s$ in P6.p, $EGF_2 = s\gamma$ in P5.p and $EGF_1 = 2\gamma^2$ in P4.p.

As before, P6.p assumes primary fate (see Table 4.1). With a similar reasoning as the one given before, this can only happen if the control parameters corresponding to P6.p satisfy the condition $c_3 = s + q_3 \geq 0$.

4.2.3 Priors

In order to define the priors, we need to take into account their range (see Table 4.2) and the constraints imposed on them by the data (see Subsection 4.2.2) and by the model (see Table 3.2). The priors are chosen so that they are fairly non informative but still reflecting our knowledge about their ranges.

For bounded parameters $m_{11}, m_{12}, m_{21}, \gamma$ and l_d we assign uniform priors on their range.

Positive non-bounded parameters $1/\tau, H, H_1, H_2, s, l$ and s_E are assigned Gamma priors $\text{Ga}(\kappa, \beta)$ where κ and β are the shape and scale of the distribution, respectively. Note that the probability density function of $U \sim \text{Ga}(\kappa, \beta)$ is given by

$$\pi_U(u) = \frac{\beta^\kappa}{\Gamma(\kappa)} u^{\kappa-1} e^{-\beta u}, \tag{4.14}$$

where Γ is the gamma function. We choose $\kappa = 2$, $\beta = 2$, and in this case $\Gamma(\kappa) = \Gamma(2) = (2-1)! = 1$. For the positive and non-bounded parameter M, since it needs to be bigger than $M_{min}(H)$ (see Equation 4.8), we impose $M \sim M_{min}(H) + \text{Ga}(2, 2)$.

On the other hand, M_1, M_2, q_1 and q_2 are assigned Normal priors $\mathcal{N}(\mu, \sigma)$ where μ is the mean of the distribution and σ is the standard deviation. Note that the probability density function of $U \sim \mathcal{N}(\mu, \sigma)$ is given by

$$\pi_U(u) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(u-\mu)^2}{2\sigma^2}}. \tag{4.15}$$

We choose $\mu = 0$ and $\sigma = 2$.

For the parameter α we would like to favour smaller values, so we choose an exponential prior $\text{Exp}(\lambda)$, where the probability density function of $U \sim \text{Exp}(\lambda)$ is

$$\pi_U(u) = \frac{1}{\lambda} e^{-\frac{u}{\lambda}}. \tag{4.16}$$

We choose $\lambda = 1$.

The parameter q_3 is a negative non-bounded parameter, so we impose $q_3 \sim -\text{Ga}(2, 2)$.

And finally, since we would like to keep the noise level relatively low, we give σ_{dif} a flat prior on $[0.05, 1]$, i.e. a uniform distribution $\mathcal{U}([0.05, 1])$.

Table 4.3 contains a list of the priors for each parameter.

Table 4.3: *Table of priors for the parameters of the model. $\mathcal{U}([a,b])$ corresponds to the Uniform distribution in $[a,b]$. $Ga(\kappa,\beta)$ corresponds to the Gamma distribution with parameters κ, β. $\mathcal{N}(\mu,\sigma)$ corresponds to the Normal distribution with mean μ and standard deviation σ. $Exp(\lambda)$ corresponds to the Exponential distribution with parameter λ.*

Parameters	**Priors**	
	Type I	**Type II**
$1/\tau$	$\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
H	$\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
M	$M_{min}+\mathrm{Ga}(2,2)$	$M_{min}+\mathrm{Ga}(2,2)$
m_{11}, m_{12}	$\mathcal{U}([-1,0])$	$\mathcal{U}([0,1])$
m_{21}	$\mathcal{U}([-1,1])$	$\mathcal{U}([-1,1])$
q_1, q_2	$\mathcal{N}(0,4)$	$\mathcal{N}(0,4)$
q_3	$-\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
γ	$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$
H_1, H_2	$\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
M_1, M_2	$\mathcal{N}(0,2)$	$\mathcal{N}(0,2)$
s, l	$\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
α	$\mathrm{Exp}(1)$	$\mathrm{Exp}(1)$
l_d	$\mathcal{U}([0,1])$	$\mathcal{U}([0,1])$
s_E	$\mathrm{Ga}(2,2)$	$\mathrm{Ga}(2,2)$
σ_{dif}	$\mathcal{U}([0.05,1])$	$\mathcal{U}([0.05,1])$

4.2.4 Distance

In order to approximate the posterior distribution of the parameters, as explained in Section 2.4 of Chapter 2, the ABC SMC algorithm computes a sequence of intermediate distributions that are obtained by comparing the simulated data and the experimental data. This comparison is made by means of a distance function d that measures the level of similarity between two data sets.

As explained in the introduction to this chapter, Table 4.1 provides 99 data points which correspond to the proportion of times that VPC c ($c = 1,2,3$ for P4.p, P5.p and P6.p respectively), adopted fate f ($f = 1,2,3$ for $1^\circ, 2^\circ$ and 3° respectively) in experiment e ($e = 1,2,\ldots,13$ for the corresponding experiment in

Table 4.1). Each data point will be represented by $p_{e,f,c}$.

Given a parameter vector $\boldsymbol{\theta}$, we can simulate each data point using our model as explained in Section 3.4 of Chapter 3, obtaining the corresponding $p^{sim}_{e,f,c}(\boldsymbol{\theta})$.

If $\boldsymbol{X}_0^{\Omega}$ is a subset of data corresponding to experiments $\Omega_{exp} = \{e_1, \ldots, e_{N_{exp}}\}$ in Table 4.1, we define the distance between $\boldsymbol{X}_0$ and the corresponding simulation $\boldsymbol{X}^{\Omega}(\boldsymbol{\theta})$ as

$$d(\boldsymbol{X}_0^{\Omega}, \boldsymbol{X}^{\Omega}(\boldsymbol{\theta})) = \frac{1}{N_{exp}} \sum_{e \in \Omega_{exp}} \sum_{f=1}^{3} \sum_{c=1}^{3} \left| p_{e,f,c} - p^{sim}_{e,f,c}(\boldsymbol{\theta}) \right| + \frac{1}{N_{exp}} \sum_{e \in \Omega_{exp}} \sum_{c=1}^{3} \left| p^{sim}_{e,4,c}(\boldsymbol{\theta}) \right|, \tag{4.17}$$

where N_{exp} is the number of experiments in X_0, i.e. the size of Ω, and $p^{sim}_{e,4,c}$ is the proportion of times that our model could not assign a fate to cell c in experiment e. The distance function penalises parameter values for which our model cannot assign a fate, since we would like to avoid these parameter values.

The distance defined in Equation 4.17 is similar to an L^1 norm since it is linear on the differences. Similarly, a distance based on the L^2 norm could be defined, by squaring the differences and taking the squared root of the sum. There is not much literature comparing the efficiency of different distances. However, Mckinley et al. (2009) find little difference between these two distances in their fitting results so we decide to use the distance defined in Equation 4.17.

Further work would be needed to check whether a better distance could be defined. We will expand on this idea in Section 4.4.

4.2.5 Sensitivity analysis

We perform sensitivity analysis to study how the results of our model are affected by changes in the parameters. This can helps us to find the parameters that our model is more sensitive to.

Let us denote by $\boldsymbol{X}^e(\boldsymbol{\theta})$ the outcome of our model for experiment e with $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_{19})$ as parameter. Each θ_i corresponds to one of the parameters in Table 4.3, therefore $i = 1, \ldots, 19$. In order to measure the sensitivity of the model for that experiment with respect to θ_i, one can measure how different is $\boldsymbol{X}^e(\boldsymbol{\theta})$ from $\boldsymbol{X}^e(\rho_{i,\delta}(\boldsymbol{\theta}))$ where the function $\rho_{i,\delta}$ has the form

$$\rho_{i,\delta}(\boldsymbol{\theta}) = (\theta_1, \theta_2, \ldots, \delta\theta_i, \ldots, \theta_{19}), \tag{4.18}$$

and $\delta > 0$. In other words, the function $\rho_{i,\delta}$ perturbs the i-th component of the parameter vector $\boldsymbol{\theta}$ by a factor of δ.

To perform the sensitivity analysis, we take twenty different parameter vectors $\boldsymbol{\theta}^1, \ldots, \boldsymbol{\theta}^{20}$ and compute $d^{e,j} = d\left(\boldsymbol{X}_0^e, \boldsymbol{X}^e(\boldsymbol{\theta}^j)\right)$, the distance (as defined in subsection 4.2.4) between $\boldsymbol{X}_0^e$, the experimental data corresponding to experiment e, and $\boldsymbol{X}^e\left(\boldsymbol{\theta}^j\right)$, the data corresponding to the same experiment simulated with the parameter vector $\boldsymbol{\theta}^j$, for $j = 1, \ldots, 20$.

For every parameter θ_i^j (the i-th component of the parameter vector $\boldsymbol{\theta}^j$), we also compute the distance $d_{i,\delta}^{e,j} = d\left(\boldsymbol{X}_0^e, \boldsymbol{X}^e(\rho_{i,\delta}(\boldsymbol{\theta}^j))\right)$, for $\delta \in \mathcal{A} = \{0.5, 0.8, 0.9, 1.1, 1.2, 1.5\}$.

We define the sensitivity of the model with respect to the i-th parameter and experiment e as the value

$$S_i^e = \sum_{j=1}^{20} \sum_{\delta \in \mathcal{A}} |d^{e,j} - d_{i,\delta}^{e,j}|. \tag{4.19}$$

If we would like to measure the sensitivity of the model with respect to the i-th parameter in several experiments, $\Omega = \{e_1, \ldots, e_k\}$, we take the value

$$S_i^\Omega = \sum_{e \in \Omega} \sum_{j=1}^{20} \sum_{\delta \in \mathcal{A}} |d^{e,j} - d_{i,\delta}^{e,j}|. \tag{4.20}$$

We repeat this analysis for the Type I and Type II models, considering $\Omega_1 = \{1, 2, 3, 4, 5, 6, 7\}$ and $\Omega_2 = \{1, 2, \ldots, 12, 13\}$. We obtain Figures 4.3–4.6.

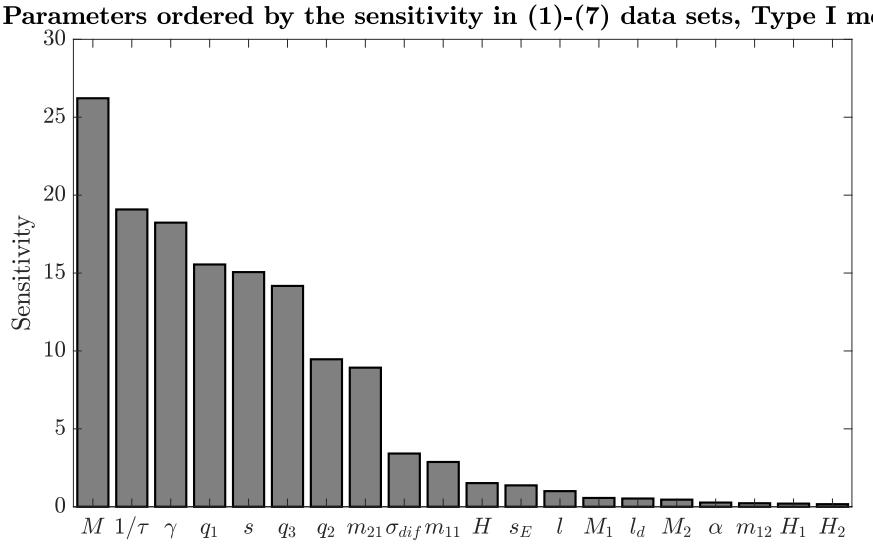


Figure 4.3: *Sensitivity plot of the Type I model with respect to each parameter and data sets (1)-(7) in Table 4.1. We can see that the model is especially sensitive with respect to 8 parameters, M being the most important one.*

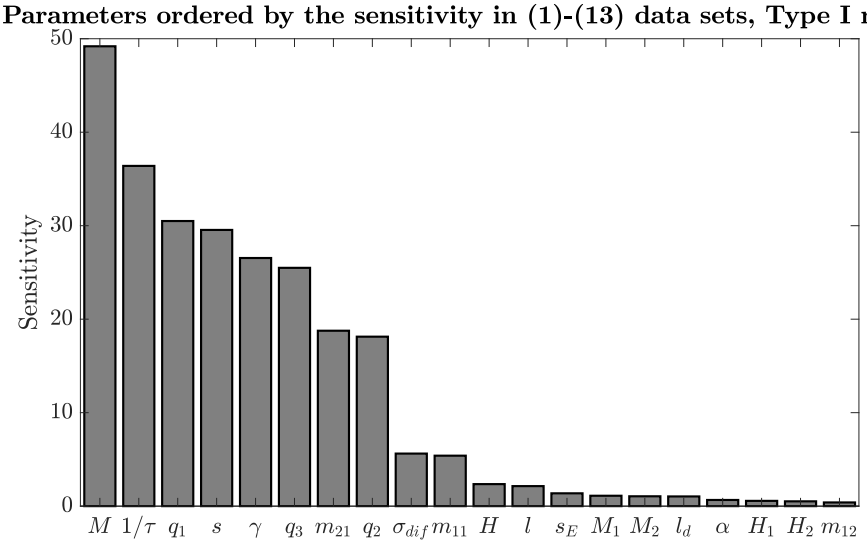


Figure 4.4: *Sensitivity plot of the Type I model with respect to each parameter and data sets (1)-(13) in Table 4.1. We can see that the model is especially sensitive with respect to the same 8 parameters as in Figure 4.3, M being again the most important parameter.*

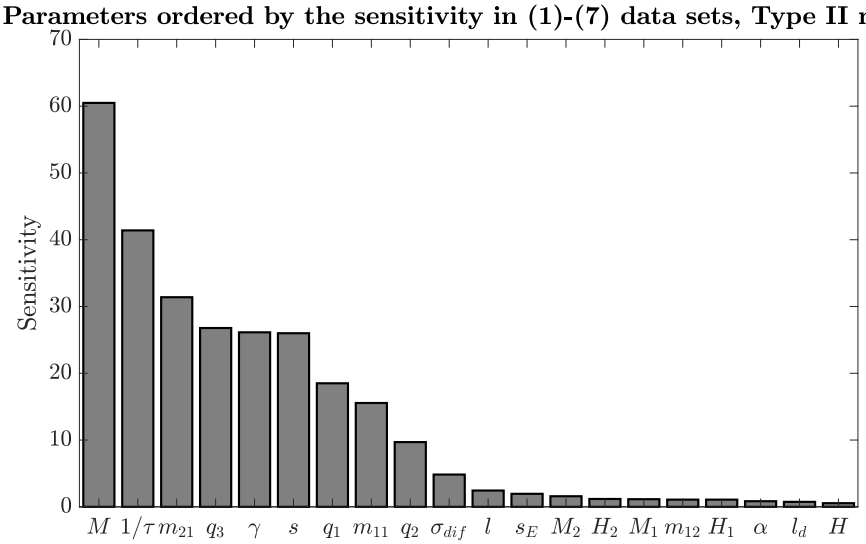


Figure 4.5: *Sensitivity plot of the Type II model with respect to each parameter and data sets (1)-(7) in Table 4.1. We can see that the model is especially sensitive with respect to 8 parameters, M being the most important one. In this case, m_{11} is in the top 8 while l is not, in comparison with Type I model top 8 parameters.*

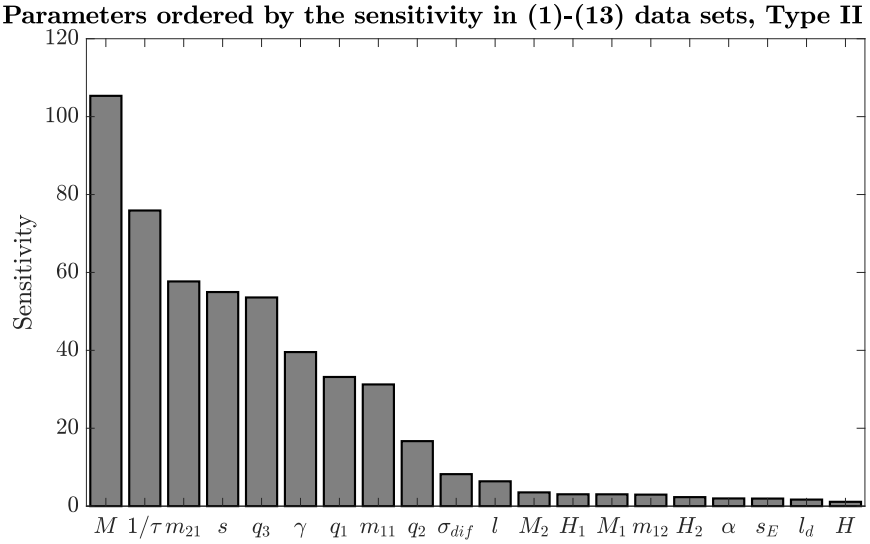


Figure 4.6: *Sensitivity plot of the Type II model with respect to each parameter and data sets (1)-(13) in Table 4.1. We can see that the model is especially sensitive with respect to the same 8 parameters as in Figure 4.5, M being again the most important parameter.*

As a result of the sensitivity analysis, we decide to fit the parameters M, $1/\tau$, q_1, s, γ, q_3, m_{21} and q_2 for the Type I model. We also add σ_{dif} and m_{11} to the list of parameters to fit because if we look at the sensitivity for each one of the 20 particles considered in the study, we find that they also account for an important proportion of variability in the results.

On the other hand, we decide to consider M, $1/\tau$, m_{21}, q_3, γ, s, q_1, m_{11} and q_2 for the fitting of the Type II model as they are the 8 parameters for which the model is more sensitive. We also add σ_{dif}, l and s_E because they also account for an important proportion of variability in the results when looking at the sensitivity for each one of the 20 particles considered in the study.

The rest of the parameters are fixed to take the values obtained from the mean of the 20 particles sampled in each case (see Table 4.4).

4.2.6 Sequence of thresholds and number of particles

As explained in Section 2.4 of Chapter 2, the ABC SMC algorithm approximates the posterior distribution of the parameters by sequentially sampling from the distributions in Equation 4.5. For this, one needs to choose a decreasing sequence of tolerances $\varepsilon_1 > \varepsilon_2 > \ldots \varepsilon_T$ which should lead to the approximation of the posterior p_{ε_T}.

Table 4.4: *Tables of fixed parameters and their corresponding values in Type I and Type II models.*

Type I model	
Parameter	**Fixed value**
H	25
m_{12}	-0.3
H_1	5
H_2	4
M_1	-0.14
M_2	0.25
l	2.63
α	0.76
l_d	1

Type II model	
Parameter	**Fixed value**
H	25
m_{12}	0.23
H_1	4
H_2	3
M_1	-0.1
M_2	0.15
α	0.8
l_d	1

The most common approach is to choose ε_{t+1} to be the α quantile of the population of particles obtained at time t (Beaumont et al. 2009, Liepe et al. 2014), taking into account a trade off between acceptance rate of the particles and the succesful convergence of the method (Silk et al. 2012). We use the $\alpha = 0.03$ quantile.

An advantage of ABC SMC is that it can be easily parallelised since the N particles corresponding to a step t are sampled independently. With that in mind, taking into account the number of parameters to be fitted and the computational time of the simulations, we choose to sample $N = 10^4$ particles at each step t of the algorithm, and parallelise the computations taking advantage of 100 cores which compute 100 subsets of 100 particles each, significantly reducing the computation time of the algorithm.

4.2.7 Perturbation kernel

The choice of perturbation kernel is also very important as it can speed up the convergence of the method (Filippi et al. 2013) if it samples from the *right* regions of parameter space.

For simplicity, let us denote by $\boldsymbol{X}_0^{\Omega} = \boldsymbol{X}_0$ a set of experimental results for experiments $\Omega = \{e_1, \ldots, e_{N_{exp}}\}$. We denote the i-th particle obtained at step $t-1$ as $\boldsymbol{\theta}^{(i,t-1)}$, its corresponding weight as $w^{(i,t-1)}$ and the data simulated with $\boldsymbol{\theta}^{(i,t-1)}$,

for the same experiments as in $\boldsymbol{X}_0$, as $\boldsymbol{X}(\boldsymbol{\theta}^{(i,t-1)})$. It holds that

$$d(\boldsymbol{X}_0, \boldsymbol{X}(\boldsymbol{\theta}^{(i,t-1)})) \leq \varepsilon_{t-1} \quad \forall 1 \leq i \leq N, \tag{4.21}$$

where N is the number of sampled particles at each step.

Recalling Algorithm 6 in Chapter 2, the particles at step t are computed by randomly choosing a particle $\boldsymbol{\theta}^{(*,t-1)}$ from the previous population $\{\boldsymbol{\theta}^{(i,t-1)}\}_{i=1}^{N}$ with weights $\{\omega^{(i,t-1)}\}_{i=1}^{N}$, perturbing it with a chosen kernel $K_t(\cdot|\boldsymbol{\theta}^{(*,t-1)})$ and checking that the candidate parameter vector $\boldsymbol{\theta}^{**}$ obtained from that perturbation satisfies $d(\boldsymbol{X}_0, \boldsymbol{X}(\boldsymbol{\theta}^{**})) < \varepsilon_t$.

Many types of perturbation kernels $K_t(\cdot|\cdot)$ can be used. The most simple choice would be the component-wise perturbation kernel, in which the parameter vector $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_m)$ (m being the number of parameters to be estimated) is perturbed component-wise. In other words, each parameter θ_j is perturbed according to a Gaussian distribution $K_t(\cdot|\theta_j) = K(\cdot|\theta_j) = \mathcal{N}(\theta_j, \sigma_j)$ or Uniform distribution $\mathcal{U}([\theta_j - \sigma_j, \theta_j + \sigma_j])$. Note that, in this case, there is a kernel distribution for each component of the parameter vector and it is independent of the other components. Moreover, the kernel distributions are the same for all t. As a result, these kernels are not very efficient since they do not take into account the fact that some parameters can be highly correlated.

Following the study in Filippi et al. (2013) we decide to implement the **multivariate normal kernel with optimal covariance matrix (OLCM)**. This perturbation kernel considers a multivariate normal distribution around each particle (assessing the correlation between the parameters) and, moreover, the covariance matrix will differ from particle to particle, taking into account the structure of the *good* particles sampled. Let us define the following set

$$\left\{\left(\tilde{\boldsymbol{\theta}}^{(k,t-1)}, \tilde{\omega}^{(k,t-1)}\right)\right\}_{k=1}^{N_0} = \left\{\left(\boldsymbol{\theta}^{(i,t-1)}, \frac{\omega^{(i,t-1)}}{\bar{\omega}}\right) : d(\boldsymbol{X}_0, \boldsymbol{X}(\boldsymbol{\theta}^{(i,t-1)})) \leq \varepsilon_t, 1 \leq i \leq N\right\}, \tag{4.22}$$

where $\bar{\omega}$ is a normalising constant such that $\sum_{k=1}^{N_0} \tilde{\omega}^{(k,t-1)} = 1$. The set in Equation (4.22) is the set of particles of the population obtained at time $t-1$ for which the simulated data is closer to the experimental data than the current threshold ε_t, i.e. the set of *good* particles at time $t-1$. The weights $\bar{\omega}$ are obtained by normalising this subset of N_0 particles.

In this approach, the perturbation kernel follows a normal distribution, $K_t(\cdot \mid$

$\boldsymbol{\theta}^{(*,t-1)}) \sim \mathcal{N}(\boldsymbol{\theta}^{(*,t-1)}, \boldsymbol{\Sigma}^{(t)}_{\boldsymbol{\theta}^{(*,t-1)}})$, which is centered at the particle that will be perturbed $\boldsymbol{\theta}^{(*,t-1)}$ with covariance matrix also dependent on the particle value. Filippi et al. (2013) propose the optimal covariance matrix $\boldsymbol{\Sigma}^{(t)}_{\boldsymbol{\theta}^{(*,t-1)}}$ to be equal to the covariance of the set of particles from step $t-1$ whose distance is smaller than the current threshold ε_t (i.e. particles in the set described in Equation 4.22) plus a bias term related to the discrepancy between the mean of the particles in such population and the particle of interest $\boldsymbol{\theta}^{(*,t-1)}$:

$$\boldsymbol{\Sigma}^{(t)}_{\boldsymbol{\theta}^{(*,t-1)}} \approx \sum_{k=1}^{N_0} \tilde{\omega}^{(k,t-1)} \left(\tilde{\boldsymbol{\theta}}^{(k,t-1)} - \tilde{\boldsymbol{m}}\right) \left(\tilde{\boldsymbol{\theta}}^{(k,t-1)} - \tilde{\boldsymbol{m}}\right)^T + \left(\tilde{\boldsymbol{m}} - \boldsymbol{\theta}^{(*,t-1)}\right) \left(\tilde{\boldsymbol{m}} - \boldsymbol{\theta}^{(*,t-1)}\right)^T, \quad (4.23)$$

where $\tilde{\boldsymbol{m}} = \sum_{k=1}^{N_0} \tilde{\omega}^{(k,t-1)} \tilde{\boldsymbol{\theta}}^{(k,t-1)}$ is the mean of the population of particles described in Equation 4.22. We therefore implement these multivariate normal perturbation kernels $K_t(\cdot \mid \boldsymbol{\theta}^{(*,t-1)}) \sim \mathcal{N}(\boldsymbol{\theta}^{(*,t-1)}, \boldsymbol{\Sigma}^{(t)}_{\boldsymbol{\theta}^{(*,t-1)}})$ with optimal covariance matrix given in Equation 4.23.

4.3 Results

Here we present the results of our parameter fitting. On what follows, the approximation of posterior distribution of the parameter vector $\boldsymbol{\theta}$ and given data $\boldsymbol{X}_0$ obtained from ABC SMC algorithm with threshold ε will be denoted by $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0; MT_1)$ if we used Type I model, and $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0; MT_2)$ if we used Type II model. We will denote the vector of experimental results for experiments $\Omega = \{e_1, \ldots, e_{Nexp}\}$ by $\boldsymbol{X}_0^\Omega$, and the experimental results for experiment e_i by $\boldsymbol{X}_0^{e_i}$. Similarly, we will denote the vector of simulated results for experiments $\Omega = \{e_1, \ldots, e_{N_{exp}}\}$ with parameter vector $\boldsymbol{\theta}$ by $\boldsymbol{X}_1^\Omega(\boldsymbol{\theta})$ if they were obtained with Type I model and by $\boldsymbol{X}_2^\Omega(\boldsymbol{\theta})$ if they were obtained with Type II model. If, instead, the results were simulated for one experiment e_i, they will be denoted by $\boldsymbol{X}_1^{e_i}(\boldsymbol{\theta})$ and $\boldsymbol{X}_2^{e_i}(\boldsymbol{\theta})$ accordingly.

4.3.1 Reproducing experiments (1) to (7)

Following the pipeline described in the previous section, we first estimate the parameters that reproduce the set of experimental results for the first seven experiments in Table 4.1, denoted by $\boldsymbol{X}_0^{\Omega_{1-7}}$. As described in Subsection 4.2.5, we fix some of the parameters and fit the rest of them using ABC SMC. We can show that both

Type I and Type II model can reproduce the experimental data $\boldsymbol{X}_0^{\Omega_{1-7}}$ and that ABC SMC can approximate the posterior distribution of the parameters.

4.3.1.1 Type I model

Starting from the prior distributions described in Subsection 4.2.3, we use the ABC SMC algorithm with $N = 10^4$ and first threshold $\varepsilon_1 = 4$ to sample from the approximation of the posterior distribution $\pi_{\varepsilon_1}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1) = \pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, where

$$\boldsymbol{\theta} = (1/\tau, M, m_{11}, m_{21}, q_1, q_2, q_3, \gamma, s, l, s_E, \sigma_{dif})$$

and the rest of parameters take the values in Table 4.4. Figures 4.7–4.15 show the marginal distributions $\pi_4(\theta_i \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ for each of the estimated parameters and the two-dimensional scatter plots for each pair of fitted parameters. We can see that the parameters are not correlated to each other. We can already distinguish a bimodal distribution in the parameter m_{21} which, in Chapter 3, we predicted should be positive.

The sequence of thresholds is obtained from the 0.03-quantiles of the populations of particles at each step of the algorithm, as explained in Subsection 4.2.6. In this case, it is given by $\varepsilon_1 = 4 > 2.07 > 1.06 > 0.57 > 0.3 > 0.17 > 0.11 = \varepsilon_7$. As it can bee seen in Figures 4.10–4.12, the aproximations of the posterior distribution given by $\pi_{0.17}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ and $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ are very similar. Also, the data that the particles $\{\boldsymbol{\theta}^{(i,7)}\}_{i=1}^N$, which are sampled from $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, simulate approximate very well the experimental data. This can be seen in Table 4.5, where we show the mean simulated outcomes (in bold) and the experimental outcomes for each mutant. Hence, we consider that the algorithm has converged when $T = 7$ and $\varepsilon_7 = 0.11$, and that the Type I model fits the experimental data for the first seven mutants.

The mean and 95% confidence intervals of the mean for each parameter fitted, as computed from the particles obtained at final step $T = 7$ of the ABC algorithm, are given in Table 4.6.

Table 4.5: *Table of mean simulated data (in bold) obtained from the $N = 10^4$ particles sampled from $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ with the ABC SMC algorithm for Type II model, and experimental data (underneath the mean simulated data, in regular font). The small amount that is missing for the proportions of fates in experiment (7) to sum up to 100 is due to some particles producing simulations for experiment (7) for which the model cannot score the fate.*

	Experiment	VPC fates (% 1°, 2° 3°)		
		P4.p	P5.p	P6.p
	Wild-type outcomes under reduced signalling			
(1)	Wild type	**0,0.56,99.44**	**0.53,99.47,0**	**99.96,0.04,0**
		0,0,100	0,100,0	100,0,0
(2)	*let-23* mosaic	**0,0.01,99.99**	**0,99.99,0.01**	**99.96,0.04,0**
	(no EGF receptors in P5/7.p)	0,0,100	0,100,0	100,0,0
(3)	Half dose of *lin-3*	**0,0.09,99.91**	**0.08,99.91,0.01**	**99.93,0.07,0**
	(Half EGF ligand)	0,0,100	0,100,0	100,0,0
(4)	Half dose of *lin-12*	**0.02,0.11,99.89**	**5.55,94.45,0**	**99.96,0.04,0**
	(Half Notch receptor)	0,0,100	0,100,0	100,0,0
	Phenotypes under reduced Notch/excess EGF			
(5)	Notch null, and 2×WT EGF (2 ACs)	**0,0,100**	**97.07,2.93,0**	**99.96,0.04,0**
	(No Notch receptor, twice as much WT EGF ligand)	0,0,100	100,0,0	100,0,0
(6)	No Notch signalling, WT EGF	**0,0,100**	**0,0.01,99.99**	**99.96,0.04,0**
	(No Notch receptor, WT EGF ligand)	0,0,100	0,0,100	100,0,0
(7)	EGF overexpression	**0.07,47.36,52.56**	**47.37,52.63,0**	**99.86,0.14,0**
	(overexpression of EGF ligand)	18,46,36	45.5,54.5,0	96,4,0

Table 4.6: *Mean and 95% confidence interval of each parameter obtained from fitting Type I model to the data $\boldsymbol{X}_0^{\Omega_{1-7}}$.*

Parameters	**Mean ± 95% CI**
$1/\tau$	5.3022 ± 0.0143
M	4.9804 ± 0.0100
m_{11}	-0.0848 ± 0.0008
m_{21}	0.6201 ± 0.0008
q_1	-1.3583 ± 0.0054
q_2	-0.5073 ± 0.0042

Parameters	**Mean ± 95% CI**
q_3	-0.9641 ± 0.0036
γ	0.1369 ± 0.0005
s	6.9587 ± 0.0302
s_E	3.0765 ± 0.0121
σ_{dif}	0.0580 ± 0.0001

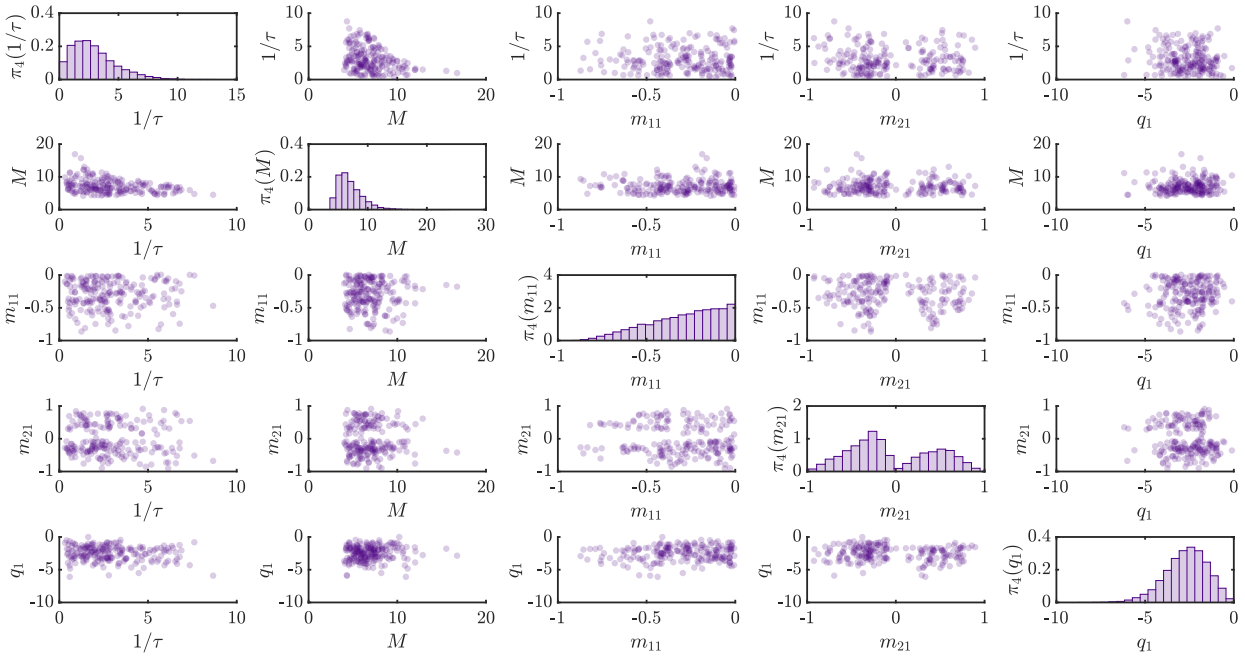


Figure 4.7: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, for the first 5 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles. We can already notice that the histogram corresponding to parameter m_{21} has more positive samples as we predicted.*

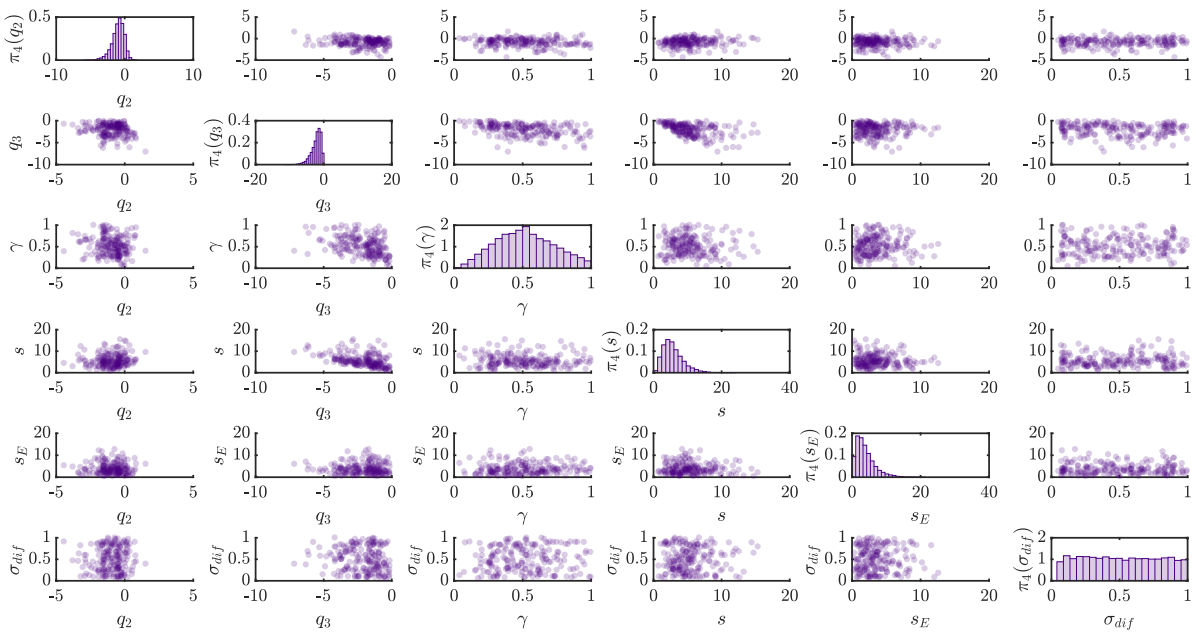


Figure 4.8: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, for the last 6 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

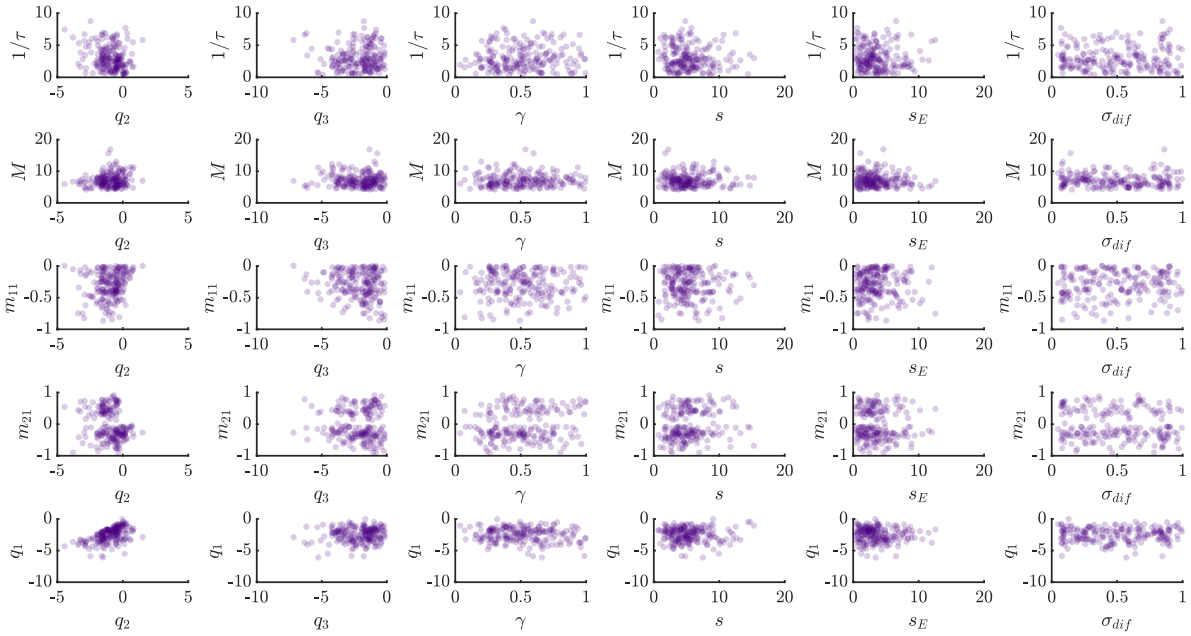


Figure 4.9: *Two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$. The scatter plots have been obtained from a subsample of 100 particles.*

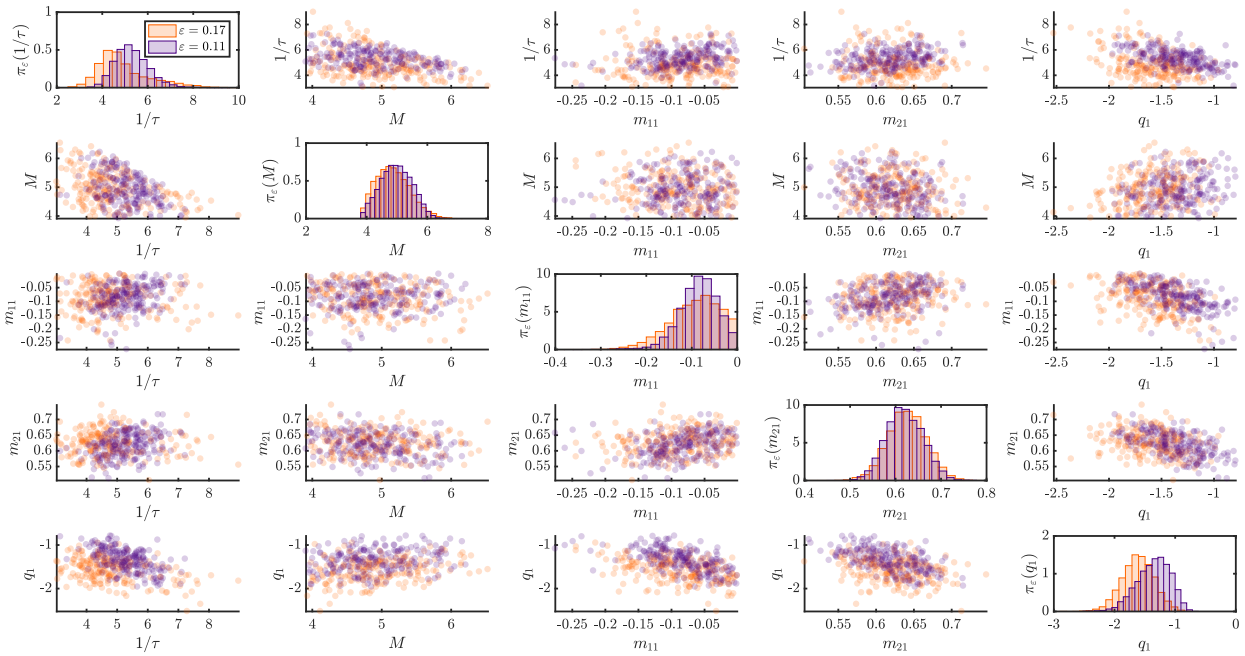


Figure 4.10: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, where $\varepsilon = 0.17$ (in orange) and $\varepsilon = 0.11$ (in purple), for the first 5 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

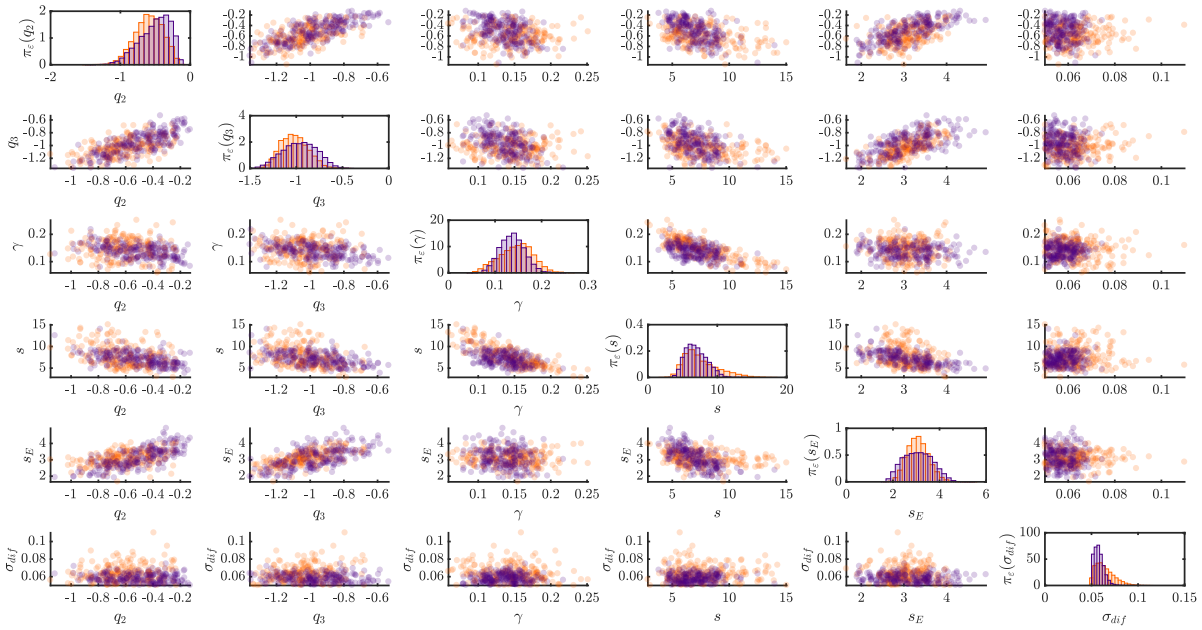


Figure 4.11: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, where $\varepsilon = 0.17$ (in orange) and $\varepsilon = 0.11$ (in purple), for the last 6 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

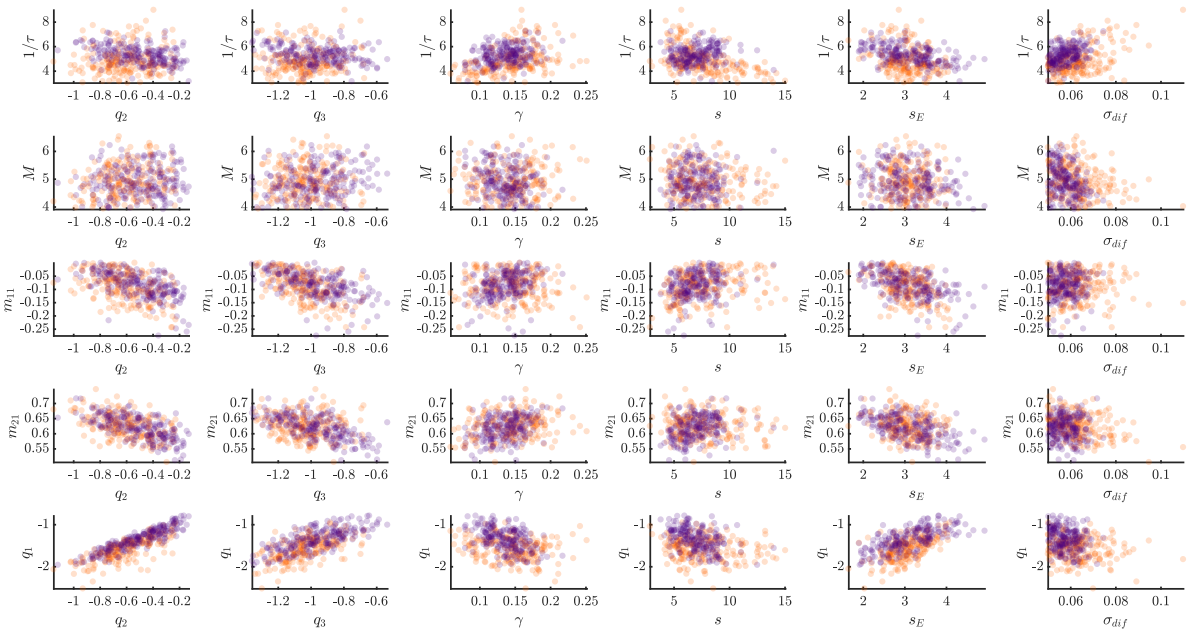


Figure 4.12: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, where $\varepsilon = 0.17$ (in orange) and $\varepsilon = 0.11$ (in purple). The scatter plots have been obtained from a subsample of 100 particles.*

4.3.1.2 Type II model

In this case, the parameters to estimate are

$$\boldsymbol{\theta} = (1/\tau, M, m_{11}, m_{21}, q_1, q_2, q_3, \gamma, s, l, s_E, \sigma_{dif})$$

and the rest of parameters take the values in Table 4.4. As before, we start from the prior distributions described in Subsection 4.2.3 and we use the ABC SMC algorithm with $N = 10^4$ and first threshold $\varepsilon_1 = 4$ to sample from the approximation of the posterior distribution $\pi_{\varepsilon_1}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2) = \pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$. Figures 4.13–4.15 show the marginal distributions $\pi_4(\theta_i \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$ and the two-dimensional scatter plots for each pair of estimated parameters. We can see that the parameters are not correlated to each other. We can also distinguish, in this case, a bimodal distribution in the parameter m_{21} which we predicted it should be positive in Chapter 3.

The sequence of thresholds is obtained from the 0.03-quantiles of the populations of particles at each step of the algorithm, as explained in Subsection 4.2.6, and is given by $\varepsilon_1 = 4 > 1.73 > 0.65 > 0.23 > 0.11 = \varepsilon_5$. The aproximations of the posterior distribution obtained with the last two thresholds, $\pi_{0.23}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$ and $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, are very similar (see Figures 4.16–4.16). Also, the data that the particles $\{\boldsymbol{\theta}^{(i,5)}\}_{i=1}^N$, which sampled from $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, approximate very well the experimental data. This can be seen in Table 4.7, where we show the mean simulated outcomes (in bold) and the experimental outcomes for each mutant. Hence, we consider that the algorithm has converged when $T = 5$ and $\varepsilon_5 = 0.11$, and that the Type II model fits the first seven experiments.

The mean and 95% confidence intervals of the mean for each parameter fitted, as computed from the particles obtained at final step $T = 5$ of the ABC algorithm, are given in Table 4.8.

Table 4.7: *Table of mean simulated data (in bold) obtained from the $N = 10^4$ particles sampled from $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$ with the ABC SMC algorithm for Type II model, and experimental data (underneath the mean simulated data, in regular font). The small amount that is missing for the proportions of fates in experiment (7) to sum up to 100 is due to some particles producing simulations for experiment (7) for which the model cannot score the fate.*

	Experiment	VPC fates (% 1°, 2° 3°)		
		P4.p	P5.p	P6.p
	Wild-type outcomes under reduced signalling			
(1)	Wild type	**0,0.13,99.87**	**0,100,0**	**99.99,0.01,0**
		0,0,100	0,100,0	100,0,0
(2)	*let-23* mosaic	**0,0.01,99.99**	**0,100,0**	**99.99,0.01,0**
	(no EGF receptors in P5/7.p)	0,0,100	0,100,0	100,0,0
(3)	Half dose of *lin-3*	**0,0.04,99.96**	**0.01,99.94,0.05**	**99.93,0.07,0**
	(Half EGF ligand)	0,0,100	0,100,0	100,0,0
(4)	Half dose of *lin-12*	**0.02,1.33,98.65**	**3.7,96.3,0**	**99.99,0.01,0**
	(Half Notch receptor)	0,0,100	0,100,0	100,0,0
	Phenotypes under reduced Notch/excess EGF			
(5)	Notch null, and 2×WT EGF (2 ACs)	**0,0,100**	**99.92,0.08,0**	**99.99,0.01,0**
	(No Notch receptor, twice as much WT EGF ligand)	0,0,100	100,0,0	100,0,0
(6)	No Notch signalling, WT EGF	**0,0,100**	**0,0,100**	**99.99,0.01,0**
	(No Notch receptor, WT EGF ligand)	0,0,100	0,0,100	100,0,0
(7)	EGF overexpression	**3.5,54.27,42.2**	**42.81,57.16,0**	**99.8,0.01,0.16**
	(overexpression of EGF ligand)	18,46,36	45.5,54.5,0	96,4,0

Table 4.8: *Mean and 95% confidence interval of each parameter obtained from fitting Type II model to the data $\boldsymbol{X}_0^{\Omega_{1-7}}$.*

Parameter	**Mean ± 95% CI**
$1/\tau$	3.2523 ± 0.0141
M	4.9211 ± 0.0129
m_{11}	0.1698 ± 0.0020
m_{21}	0.6322 ± 0.0008
q_1	-3.2001 ± 0.0119
q_2	-1.0388 ± 0.0081

Parameter	**Mean ± 95% CI**
q_3	-3.2001 ± 0.0119
γ	0.2547 ± 0.0009
s	13.7812 ± 0.0430
l	12.6540 ± 0.0508
s_E	3.1923 ± 0.0105
σ_{dif}	0.1286 ± 0.0009

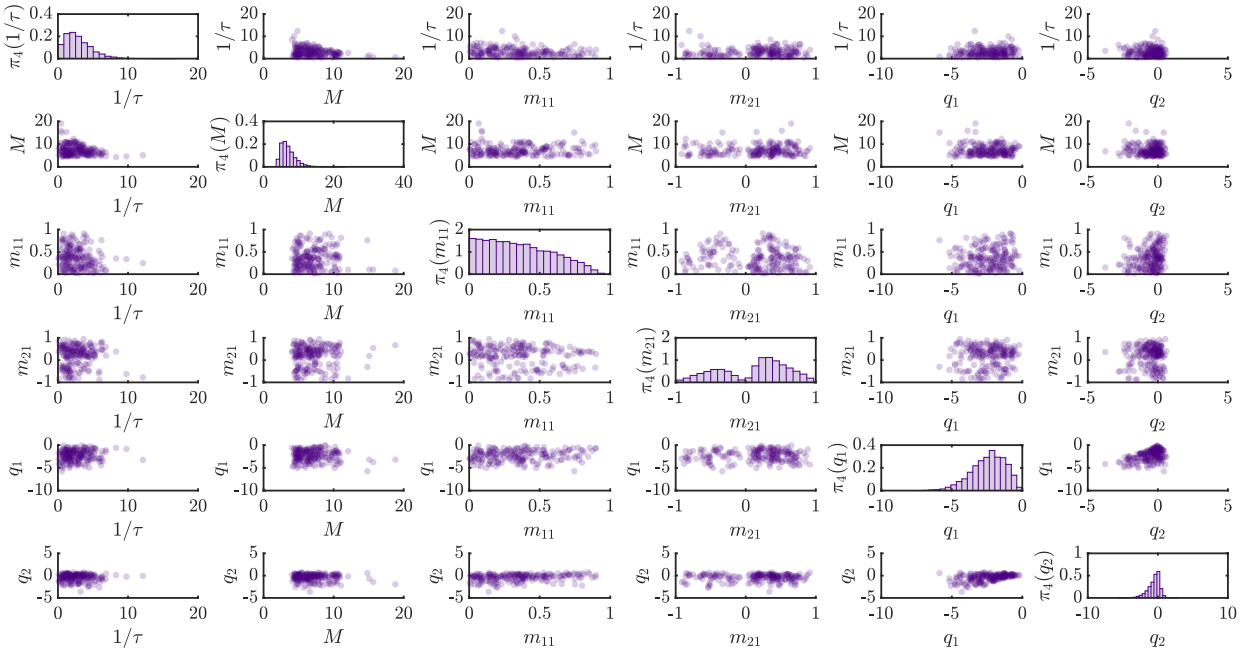


Figure 4.13: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, for the first 5 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles. We can already notice that the histogram corresponding to parameter m_{21} has more positive samples as we predicted.*

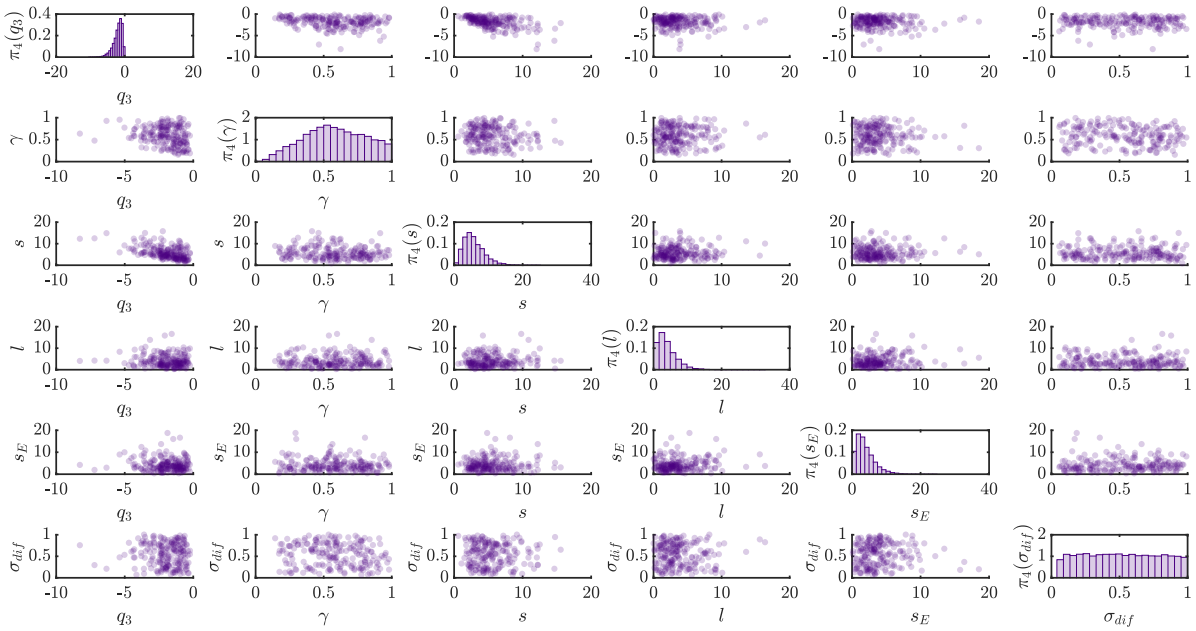


Figure 4.14: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, for the last 6 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

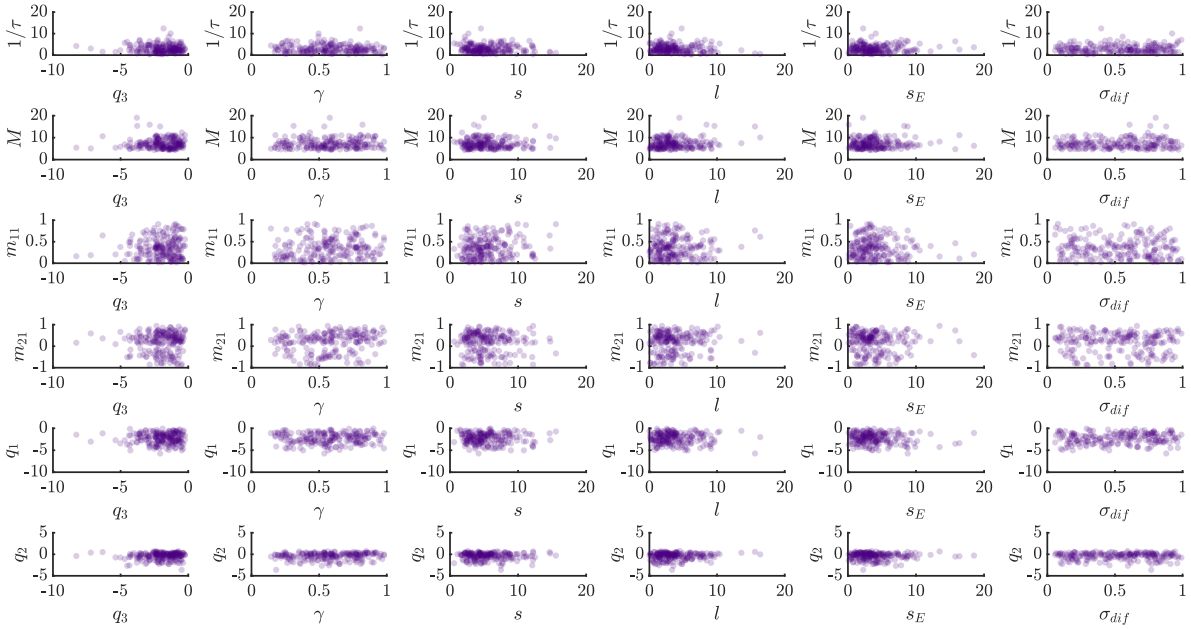


Figure 4.15: *Two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$. The scatter plots have been obtained from a subsample of 100 particles.*

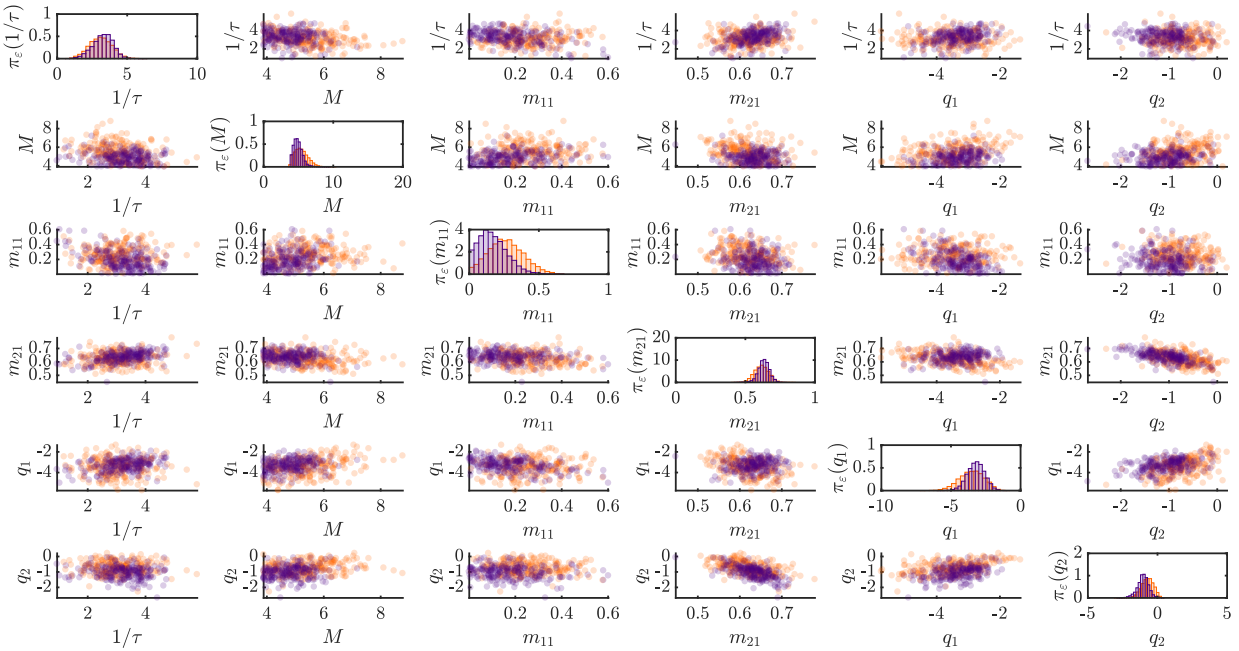


Figure 4.16: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, where $\varepsilon = 0.23$ (in orange) and $\varepsilon = 0.11$ (in purple), for the first 5 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

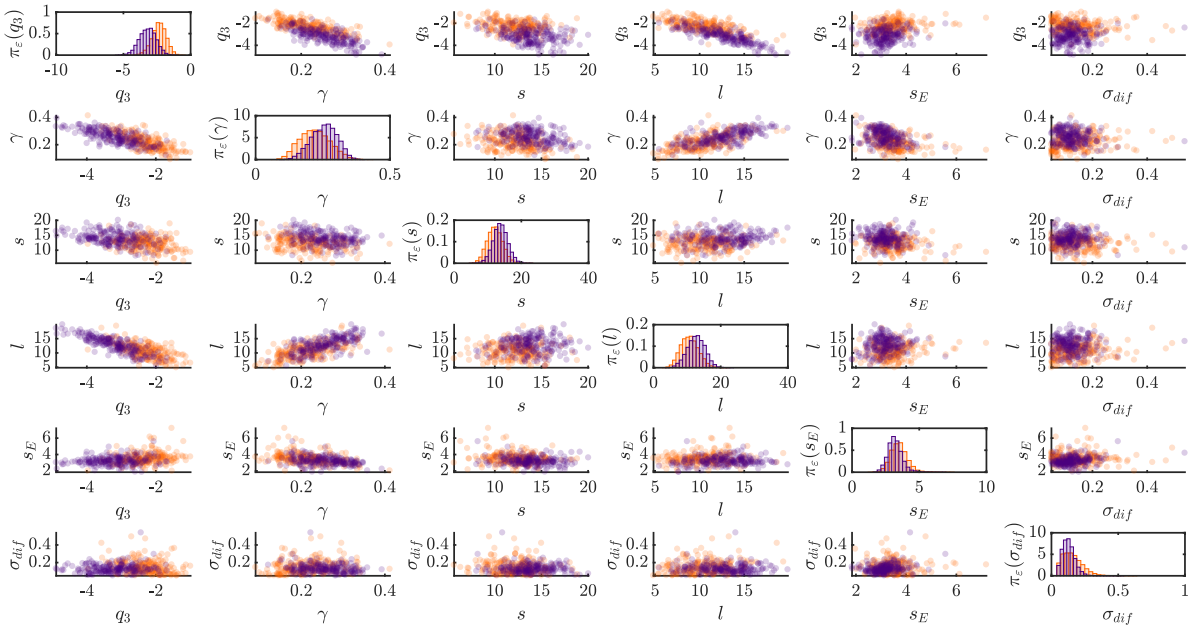


Figure 4.17: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, where $\varepsilon = 0.23$ (in orange) and $\varepsilon = 0.11$ (in purple), for the last 6 of the 12 parameters fitted. The scatter plots have been obtained from a subsample of 100 particles.*

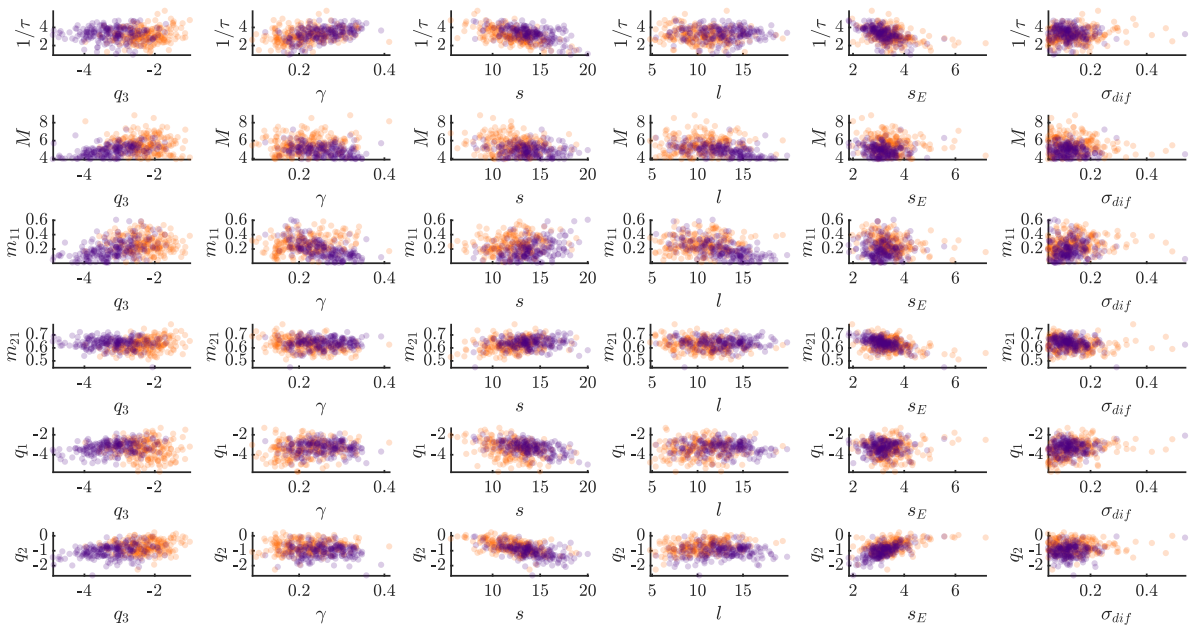


Figure 4.18: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_\varepsilon(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_2)$, where $\varepsilon = 0.23$ (in orange) and $\varepsilon = 0.11$ (in purple). The scatter plots have been obtained from a subsample of 100 particles.*

4.3.2 Fitting of experiments (1) to (8)

We proceed with the same approach for both Type I and Type II models and we obtain similar results, so here we will just show the results obtained for Type I model.

4.3.2.1 Predicting experiment (8)

First, we try to predict the outcome of experiment (8) by using the particles sampled from the approximation of the posterior distribution $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$. However, with these parameters, Table 4.9 shows that the predictions obtained for experiment (8) are not good, since P5.p would adopt 2° fate instead of 3° and P6.p would adopt 1° fate instead of 3°. In fact, the model predicts a WT pattern under AC ablation at L2 lethargus stage.

Table 4.9: *Table comparing the mean simulated data obtained with the particles sampled from $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}})$ and experiment (8) data.*

	Experiment		VPC fates (% 1°, 2° 3°)		
			P4.p	P5.p	P6.p
	Phenotypes following anchor cell ablation				
(8)	L2 lethargus	Prediction	$0, 0, 100$	$2.60, 97.28, 0.12$	$99.84, 0.16, 0$
		Exp. Data	-	$0, 0, 100$	$0, 0, 100$

Therefore we decide to perform parameter estimation with the data of the first eight experiments to estimate the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$, where

$$\boldsymbol{\theta} = (1/\tau, M, m_{11}, m_{21}, q_1, q_2, q_3, \gamma, s, l, s_E, \sigma_{dif}).$$

4.3.2.2 Fitting experiments (1) to (8)

We use ABC SMC to find the approximation of the posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$, where

$$\boldsymbol{\theta} = (1/\tau, M, m_{11}, m_{21}, q_1, q_2, q_3, \gamma, s, l, s_E, \sigma_{dif}).$$

The rest of parameters for the model are fixed and take the values in Table 4.4.

As a first threshold, we take $\varepsilon_1 = 4$, and the sequence of thresholds is obtained from the 0.03-quantiles of the populations of particles at each step of the algorithm, as explained in Subsection 4.2.6. In this case, the sequence is given by $\varepsilon_1 = 4 > 2.08 > 1.33 = \varepsilon_3$.

We find that, in this case the algorithm cannot find particles that reproduce the eight experiments at the same time. Instead, as we decrease the thresholds, the

simulated data reproduces experiments (1) to (7) better and better but stays away (in distance) from experiment (8).

In Figure 4.19 we show the evolution of the distances between the simulated data and the experimental data as we decrease the thresholds. Each point in the scatter plot comes from one of the N particles sampled, on the first subplot the particles are sampled from $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$, on the second subplot the particles are sampled from $\pi_{2.08}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$ and on the third subplot the particles are sampled from $\pi_{1.33}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$. On each plot, the x-coordinate corresponds to the distance between the data simulated with the corresponding particle and the experimental data for experiments (1) to (8), i.e. $d(\boldsymbol{X}_0^{\Omega_{1-8}}(\boldsymbol{\theta}), \boldsymbol{X}_1^{\Omega_{1-8}}(\boldsymbol{\theta}))$ where $\Omega_{1-8} = \{e_1, e_2, e_3, e_4, e_5, e_6, e_7, e_8\}$ (see Equation 4.17), and the y-coordinate corresponds to the distance between the data simulated with the corresponding particle and the experimental data for experiment (8), i.e. $d(\boldsymbol{X}_0^{e_8}(\boldsymbol{\theta}), \boldsymbol{X}_1^{e_8}(\boldsymbol{\theta}))$. We see that, even though the total distance between the simulated data for experiments (1) to (8) and the corresponding experimental data decreases, the distance to experiment (8) stays high.

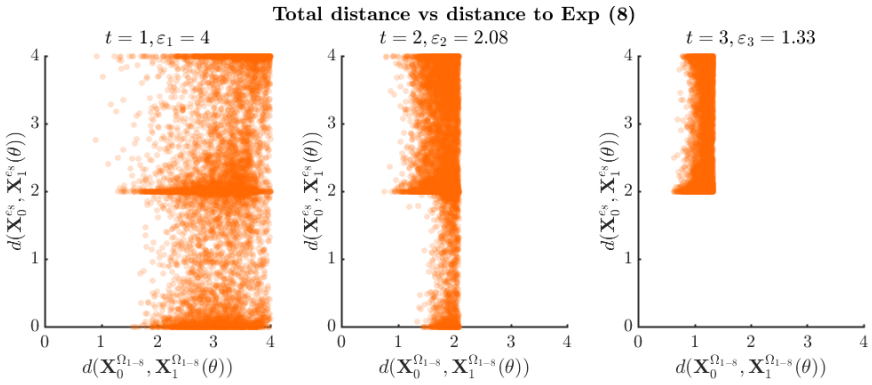


Figure 4.19: *Scatter plots showing the evolution of the distances when fitting Type I model to experiments (1) to (8).*

In Figure 4.20 we show the evolution of the distances between the simulated data and the experimental data as we decrease the thresholds, this time comparing the distances to just two experiments. As before, each point in the scatter plot comes from one of the N particles sampled, on the first subplot the particles are sampled from $\pi_4(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$, on the second subplot the particles are sampled from $\pi_{2.08}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$ and on the third subplot the particles are sampled from $\pi_{1.33}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-8}}; MT_1)$. On each plot, the x-coordinate corresponds to the distance between the data simulated with the corresponding particle and the experimental data for experiment (3), i.e. $d(\boldsymbol{X}_0^{e_3}(\boldsymbol{\theta}), \boldsymbol{X}_1^{e_3}(\boldsymbol{\theta}))$ (see Equation

4.17), and the y-coordinate corresponds to the distance between the data simulated with the corresponding particle and the experimental data for experiment (8), i.e. $d(\boldsymbol{X}_0^{e8}(\boldsymbol{\theta}), \boldsymbol{X}_1^{e8}(\boldsymbol{\theta}))$. We see that already on the first step of the algorithm, there are no particles for which simulated data reproduces both experiments. Indeed, this happens between the first four experiments and the eighth experiment, as it is showed in Figure 4.21.

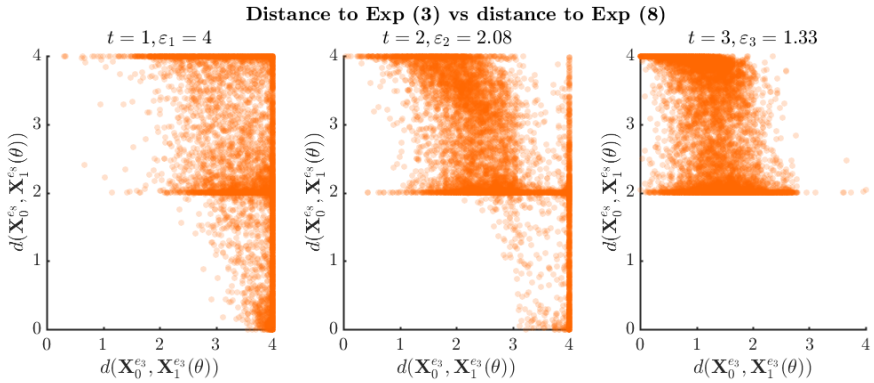


Figure 4.20: *Scatter plots showing the evolution of the distances when fitting Type I model to experiments (1) to (8).*

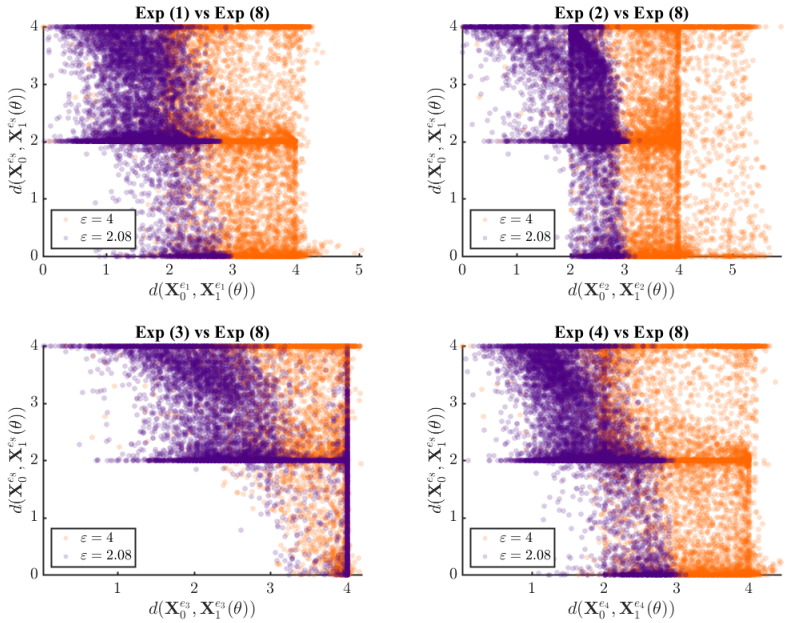


Figure 4.21: *Scatter plots showing the evolution of the distances when fitting Type I model to experiments (1) to (8).*

We then decided to check whether the model can reproduce the results of experiment (8) at all.

4.3.2.3 Fitting experiment (8)

In this case, we decide to sample directly from the posterior $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$ by using the Likelihood-free rejection sampler 1 given in Algorithm 4, since we see that, on the first step of the ABC SMC algorithm with $\varepsilon_1 = 4$, most of the particles generate simulated data at distance 0 from the experimental results for experiment (8).

Indeed, we observe that the aproximation of the posterior obtained from ABC converges to the real posterior, as one can see in Figures 4.22–4.24. The approximation of the posterior is obtained with a chain of thresholds $\varepsilon_1 = 4 < 0.26 = \varepsilon_2$, sampling $N = 10^4$ particles on each step of the algorithm.

This means that the model can reproduce experiments (1) to (7) and (8) separately, but they cannot be reproduced with the same parameter values. We will now show why we think this happens.

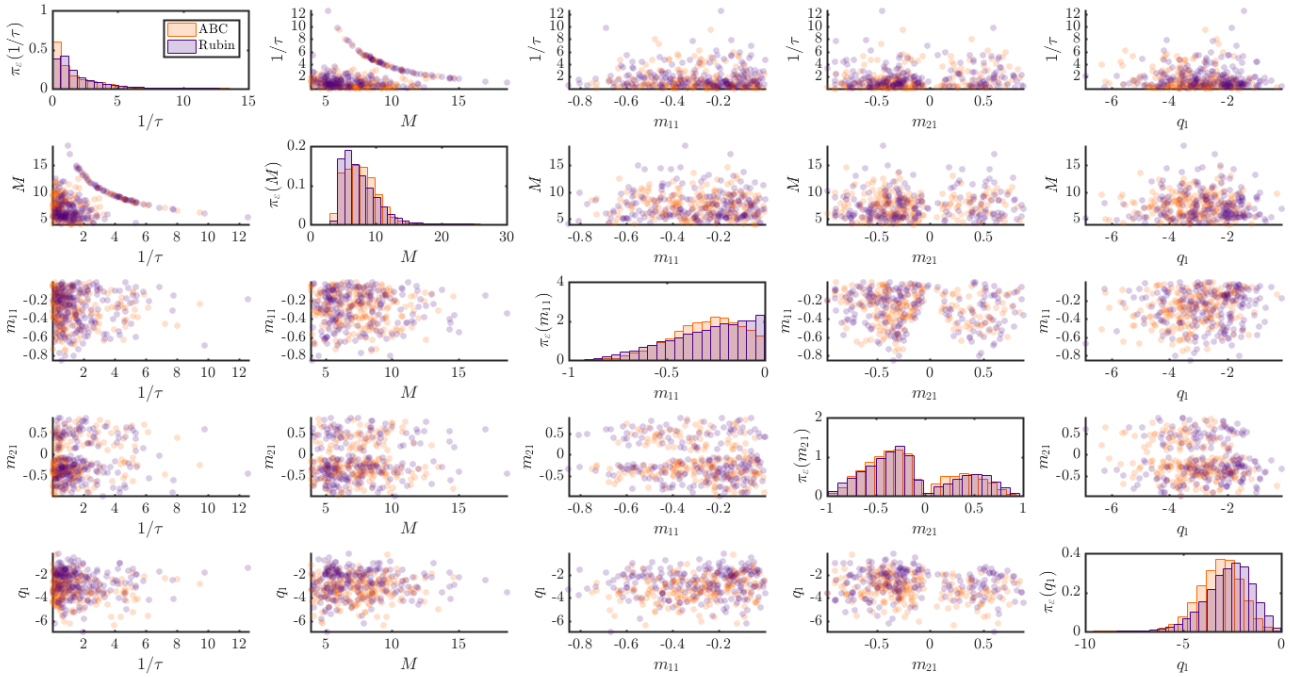


Figure 4.22: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_{0.26}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{es}; MT_1)$, where $\varepsilon = 0.26$ (in orange) and particles sampled from the real posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{es}; MT_1)$ (in purple), for the first 5 parameters fitted. The scatter plots have been obtained from a subsample of 200 particles.*

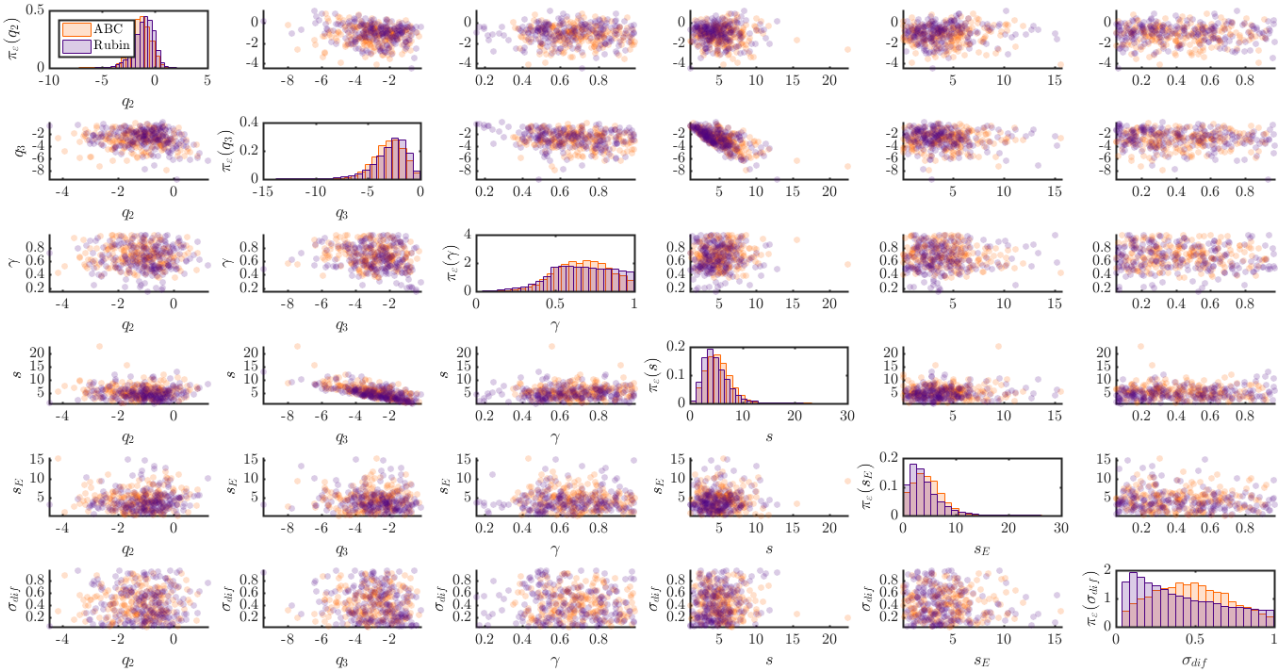


Figure 4.23: *Histograms and two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_{0.26}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{es}; MT_1)$, where $\varepsilon = 0.26$ (in orange) and particles sampled from the real posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{es}; MT_1)$ (in purple), for the last 6 to 11 parameters fitted. The scatter plots have been obtained from a subsample of 200 particles.*

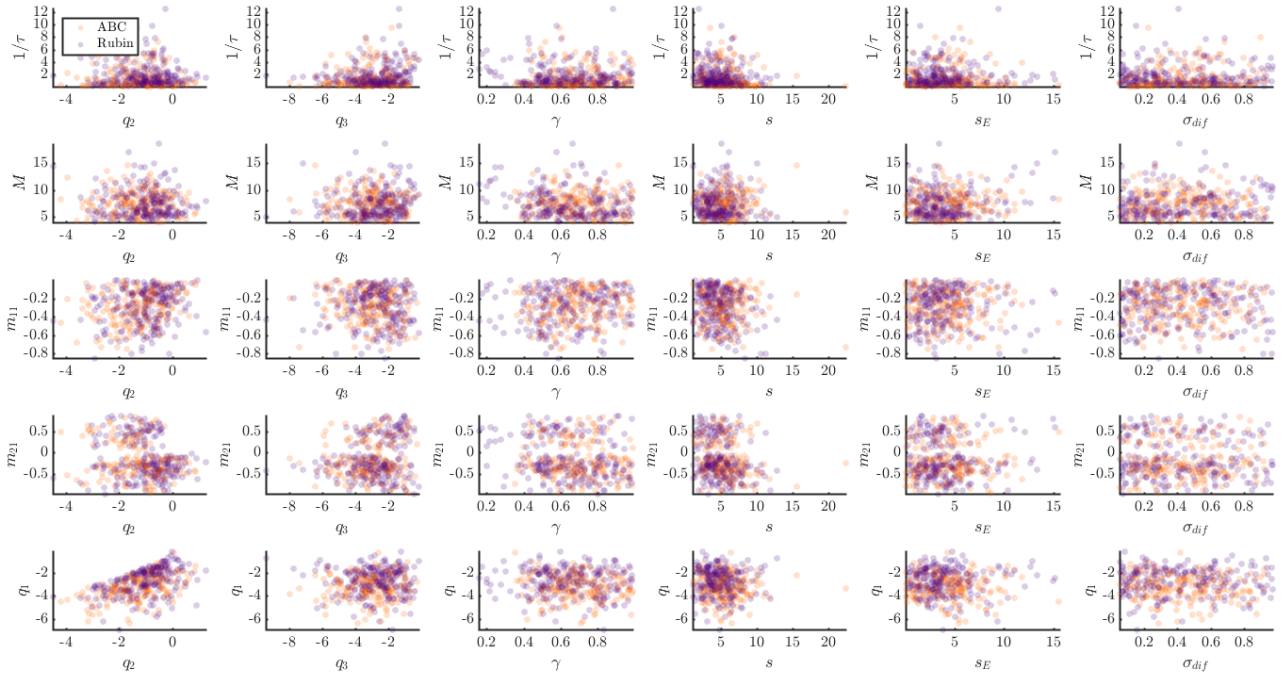


Figure 4.24: *Two dimensional scatter plots of the $N = 10^4$ particles sampled from the distribution $\pi_{0.26}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$, where $\varepsilon = 0.26$ (in orange) and particles sampled from the real posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$ (in purple). The scatter plots have been obtained from a subsample of 200 particles.*

4.3.2.4 Comparison of experiments (1) to (7) and (8)

We compare the distributions $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ and $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$, i.e. the distribution of the parameters that reproduce experiments (1) to (7) and the distribution of the parameters that reproduce experiment (8).

Figure 4.25 shows the histograms of the particles sampled from the posterior distribution $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$ (in orange) and the histograms of the particles sampled from the approximation of the posterior distribution $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ (in purple), for four parameters. We can see that the parameter values that reproduce the experiment (8) have very low probability to be sampled from the distribution $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$, and the other way around. In particular, the parameter γ needs to be very low for the Type I model to reproduce experiments (1)-(7). On the other hand, γ cannot be low for the Type I model to reproduce experiment (8).

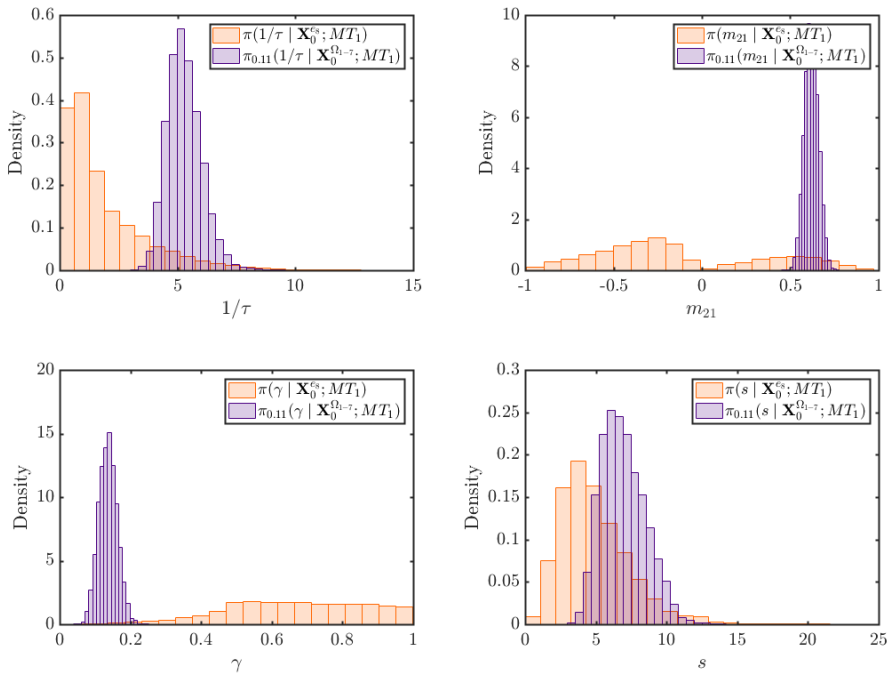


Figure 4.25: *2-dimensional scatter plots comparing the parameters that reproduce experiments (1) to (7) (in purple) and the ones that reproduce experiment (8) (in orange).*

Moreover, in Figure 4.26, we can see that if we plot the set of the particles sampled from $\pi(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{e_8}; MT_1)$ (in orange) and the set of particles sampled from the approximation of the posterior distribution $\pi_{0.11}(\boldsymbol{\theta} \mid \boldsymbol{X}_0^{\Omega_{1-7}}; MT_1)$ (in purple) in $(\gamma, 1/\tau, M)$ space, they do not intersect.

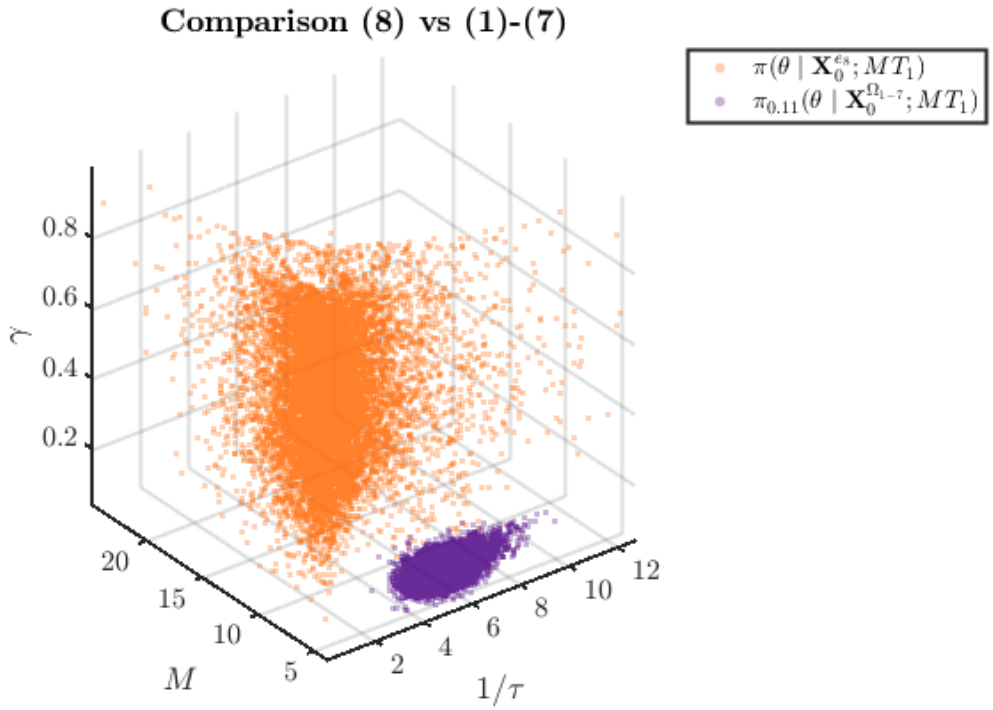


Figure 4.26: *3-dimensional scatter plot comparing the parameters that reproduce experiments (1) to (7) (in purple) and the ones that reproduce experiment (8) (in orange).*

Taking into account the parameters that have different posterior distribution when they reproduce experiments (1) to (7) or (8), they have in common that they are related to the *speed* with which the trajectories converge to the fate 1°. When we simulated experiment (8) with the particles sampled from the posterior distribution given experiments (1) to (7), the model reproduced the WT pattern, i.e. the induction in our model is very fast. On the other hand, if we simulate experiments (1) to (4) with the particles sampled from the posterior distribution given experiment (8), they are not induced and the differentiate into the $3°, 3°, 3°$ pattern, i.e. the induction is very slow. On the next section, we will discuss the results and provide some ideas about how to proceed in this case.

4.4 Discussion

In the last two chapters we have proposed a Waddington-type model for the process of vulval development in *C. elegans*. Here we review the results presented in the Chapters 3 and 4 and propose some ideas for future improvement.

4.4.1 New methodology to build quasi-potential models

In Chapter 3, we have developed a method to generate a quasi-potential model by taking advantage of the elementary catastrophes that René Thom classified in his famous theorem (see Theorem 2). Our goal was to show a method that could be used in more general and complicated cases, when the topology cannot be generated by a polynomial equation. By taking advantage of the elementary catastrophes, one can deeply study the dependence of the critical points on the parameters by means of the catastrophe manifold and bifurcation set. This, in turn, facilitates the description of the mathematical parameters by biological parameters in a non ad-hoc manner. In this case, we have consider a simple linear relation between them, but we would like to investigate more general maps between the biological signal space and the mathematical control space in the future.

4.4.2 Parameter fitting

In this chapter, we have presented the results obtained from the parameter fitting. This has been done by taking advantage of the ABC SMC algorithm, which bypasses the problem of not having a tractable likelihood. We consider that this approach improves the one given by Corson & Siggia (2012, 2017) since it accounts for stochasticity and it globally explores the parameter space, providing a real posterior for the parameters.

We have showed with this approach that our model (both Type I and Type II) can reproduce a subset of the experimental data available, and it does so very accurately. On the other hand, we have shown that we have not been able to make our model compatible with the entire data set. This could be due to three reasons: either we have not exahustively explored the parameter space and this has restricted our results, it could also be due to the nature of the data or due to the definition of our model.

We understand that there some caveats on the methodology. First of all, the convergence of the ABC SMC method highly depends on the choice of the sequence of the thresholds, as Silk et al. (2012) points out. Silk et al. (2012) assesses the difficulties of finding such a decreasing sequence of thresholds, warning that if the difference between two consecutive thresholds is not *small enough* the method could converge to local minima. The approach that they propose, however, is not applicable in this case since our model is stochastic. A possible solution would be to study the approximation of the posterior distribution obtained with different sequences of thresholds. In this regard, we have seen in Subsubsection 4.3.2.3 that

the approximation of the posterior obtained with ABC SMC method converged to the real posterior obtained with the Likelihood-free sampler 1. We would have liked to compare two approximations of the posterior distribution for each of the data sets used in the fitting but, due to lack of resources, it has not been possible.

One could also wonder if the number of particles considered is big enough to explore the entire parameter space. We have checked with twice as many particles as $N = 2 \times 10^4$, and the results were the same, so we think that this is not the main issue.

The distance function could also determine the convergence of the method. In fact, we considered to use a weighted distance to *force* the algorithm to find parameters that reproduce data set (8) when it is fitted together with experiments (1) to (7). We must say that this approach did not provide any improvement; if the weight assigned to the data set (8) was too big, the algorithm would find particles that reproduce experiment (8) but not the rest.

Regarding the nature of the data, we notice that experiments (1) to (7) are in nature different to experiments (8) to (13), since the latter ones rely on human intervention more than the former ones. We have not been able to find in the literature what the different stages in which the AC is ablated correspond to in numerical time. Therefore, it could be that there is no consistent time from animal to animal and this has not been considered in our model. We would like to obtain more specific times for the AC ablation data in order to test our model with them.

This leads us to the model we have proposed, and it opens the question of how to modify an existing model in order to reproduce a new set of data. As a result of this analysis we think that the topology that we have proposed is too restrictive. For example, the relative size of the basin of attraction of the 3° fate with respect to the 1° and 2° is fixed. This fixes the relative *speed* of the trajectory in the basin of attraction of the 3° fate with respect to the other two, and could be the reason why the two data sets are incompatible. In order for the model to reproduce experiment (8), the solution needs to stay in the basin of attraction of the fate 3° until the AC is ablated. On the other hand, for the model to reproduce experiments (1) to (4), the trajectory of P6.p needs to approach the attractor corresponding to the fate 1° fast enough so that it induces P5.p to the fate 2° before the end of the experiment. With this in mind, we tried to modify our model so that the flow defined in the basin of attraction of the 3° fate was smaller, slowing down the trajectories in that

region. In other words, we modified Equation 4.7 to be the following one:

$$\begin{cases} \frac{dx}{dt} &= \frac{1}{\tau}\varphi(y, y_*, n, \tilde{M})[\chi(y)(-4x^3 - 2ax - b) - (1 - \chi(y))x] \\ \frac{dy}{dt} &= \frac{1}{\tau}\varphi(y, y_*, n, \tilde{M})[y((y - M)^2 + c)], \end{cases} \tag{4.24}$$

where $\varphi(y, y_*, n, \tilde{M}) = 1$ if $y < y_*$ and $\varphi(y, y_*, n, \tilde{M}) = (y_*^n + \tilde{M})/(y^n + \tilde{M})$, and y_*, n and $\tilde{M}$ are parameters. Unfortunately, even with this modification we were not able to reproduce experiments (1) to (8) together, and we infer from this that the model needs to allow more bifurcations. Here we propose an idea for a new potential model for vulval development based on the compactified elliptic umbilic catastrophe.

4.4.3 Compactified elliptic umbilic catastrophe

We think that our model cannot fit the data because it is not flexible enough to allow different relative depths for the basins of attraction and bifurcations for the critical points. In order to fix this, we think that the elliptic umbilic catastrophe would be useful, in particular, a compactification of such catastrophe called **compactified elliptic umbilic**. As we will show, this compactification is formed by two terms, the elliptic umbilic catastrophe and a compactification term that controls the distances between the critical points. Here we will first introduce the elliptic umbilic catastrophe to then introduce the compactified elliptic umbilic and motivate why we think that it could reproduce the vulval development process that we studied.

4.4.3.1 Elliptic umbilic catastrophe

The elliptic umbilic is one of the elemental catastrophes in Theorem 2, in Chapter 2. It is a 3-parameter family of functions $V : \mathbb{R}^2 \times \mathbb{R}^3 \to \mathbb{R}$ given by

$$V_{u,v,w}(x, y) = x^3 - 3xy^2 + w(x^2 + y^2) + ux + vy \tag{4.25}$$

and which is the unfolding of the singularity $x^3 - 3xy^2$. This is an equivalent expression to the one given in Theorem 2.

The catastrophe manifold $\mathcal{M}$ is given by the points (x, y, u, v, w) that satisfy the two equations:

$$\frac{\partial V_{u,v,w}(x, y)}{\partial x} = 3x^2 - 3y^2 + 2wx + u = 0,$$

$$\frac{\partial V_{u,v,w}(x,y)}{\partial y} = -6xy + 2wy + v = 0.$$

$\mathcal{M}$ is a 3-dimensional manifold in $\mathbb{R}^5$ that we can write as

$$\mathcal{M} = \{(x,y,u,v,w) \in \mathbb{R}^5 | u = -3x^2 + 3y^2 - 2wx, v = 6xy - 2wy\}.$$

The intersection of $\mathcal{M}$ and the plane $\{(x,y,u,v,w)|u = u_0, v = v_0, w = w_0\}$ gives the critical points of the potential function $V_{u_0,v_0,w_0}(x,y)$.

The singularity set $\mathcal{S}$, by definition, is given by the points (x,y,u,v,w) in $\mathcal{M}$ such that (x,y) is a degenerate critical point of the potential $V_{u,v,w}(x,y)$. In other words, $\mathcal{S}$ is the set of points in $\mathcal{M}$ that satisfy:

$$\frac{\partial^2 V_{u,v,w}(x,y)}{\partial x^2}\frac{\partial^2 V_{u,v,w}(x,y)}{\partial y^2} - \frac{\partial^2 V_{u,v,w}(x,y)}{\partial xy} = 0. \tag{4.26}$$

Substituting each term by its value we obtain that, in this case, $\mathcal{S}$ is given by the set of points (x,y,u,v,w) in $\mathcal{M}$ that satisfy:

$$x^2 + y^2 = \frac{w^2}{9}, \tag{4.27}$$

which is a double cone in the (x,y,w) space (see Figure 4.27).

We can parametrise the double cone as the points $\hat{\mathcal{S}} = (\frac{w}{3}\cos(\theta), \frac{w}{3}\sin(\theta), w)$ where $w \in \mathbb{R}$ and $\theta \in [0, 2\pi]$.

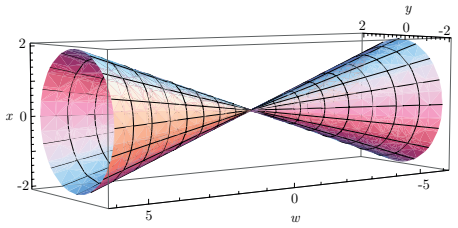


Figure 4.27: *Double cone*

The singularity set is diffeomorphic to this double cone, in particular it is

$$\mathcal{S} = \phi(\hat{\mathcal{S}}) = \left\{ \left(\frac{w}{3}\cos(\theta), \frac{w}{3}\sin(\theta), -\frac{w^2}{3}(\cos(2\theta) + 2\cos(\theta)), \frac{w^2}{3}(\sin(2\theta) - 2\sin(\theta)), w \right), \right.$$

$$\left. \theta \in [0, 2\pi] \text{ and } w \in \mathbb{R} \right\}$$

The bifurcation set $\mathcal{B}$ is the projection of the singularity set into the control space or, in other words, the image under the catastrophe map of the singularity

set:

$$\mathcal{B} = \left\{ \left(-\frac{w^2}{3}(\cos(2\theta) + 2\cos(\theta)), \frac{w^2}{3}(\sin(2\theta) - 2\sin(\theta)), w \right), \theta \in [0, 2\pi] \text{ and } w \in \mathbb{R} \right\}, \tag{4.28}$$

which is plotted in Figure 4.28.

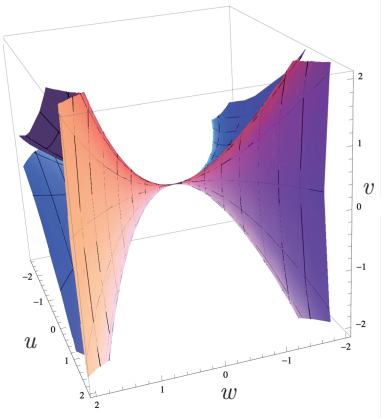


Figure 4.28: *Bifurcation set of the elliptic umbilic catastrophe.*

The constant-w cross-sections of $\mathcal{B}$ are three-cusped hypocycloids, growing parabolically. For example, we can see in Figure 4.29 the cross-section for $w = -3$.

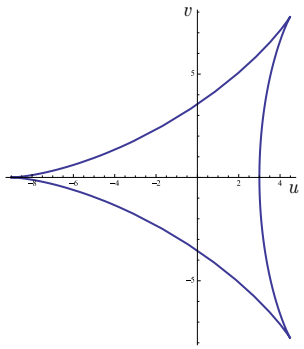


Figure 4.29: *Cross section of the bifurcation set of the elliptic umbilic catastrophe.*

Each connected component, defined by the bifurcation set $\mathcal{B}$, gives a different configuration for the equilibria. When taking parameter values on the bifurcation set, we will get a degenerate critical point obtained from the bifurcation of two or more critical points (see Figure 4.30).

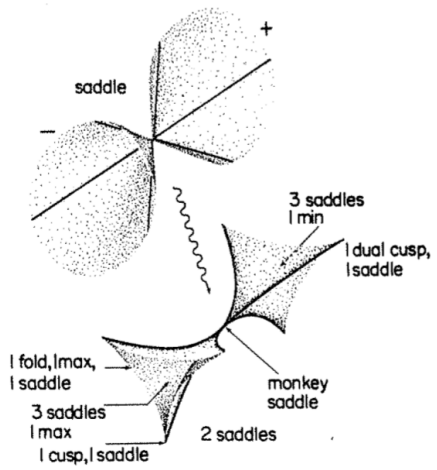


Figure 4.30: *Bifurcation set of the Elliptic Umbilic Catastrophe. Adapted from Poston & Stewart (2014).*

4.4.3.2 Compactified elliptic umbilic

The compactified elliptic umbilic is obtained by compactifying the previous catastrophe. This way, the equilibria are moved to finite positions and new dynamics appear.

$$V_{u,v,w}(x,y) = \overbrace{x^3 - 3xy^2}^{\text{Singularity}} + \overbrace{w(x^2+y^2) + ux + vy}^{\text{Unfolding}} + \overbrace{(x^4+y^4)}^{\text{Compactification}} \quad . \tag{4.29}$$

It is a three-parameter family and we are interested in studying how the change of the parameter values affects the configuration of the equilibria of the family.

Catastrophe manifold of the compactified elliptic umbilic The catastrophe manifold $\mathcal{M}$ is the set of points (x, y, u, v, w) that satisfy the following equations:

$$\begin{aligned} \tfrac{\partial V_{u,v,w}(x,y)}{\partial x} &= 3x^2 - 3y^2 + 2wx + u + 4x^3 = 0, \\ \tfrac{\partial V_{u,v,w}(x,y)}{\partial y} &= -6xy + 2wy + v + 4y^3 = 0. \end{aligned}$$

$\mathcal{M}$ is, again, a three dimensional manifold. We can use (x, y, w) as a chart on $\mathcal{M}$ where a general point of the catastrophe manifold is given by

$$(x,y,w) \overset{\phi}{\mapsto} (x, y, -3x^2 + 3y^2 - 2wx - 4x^3, 6xy - 2wy - 4y^3, w) = (x, y, u, v, w) \in \mathcal{M},$$

i.e.

$$\mathcal{M} = \{(x, y, u, v, w) \in \mathbb{R}^5 \mid u = -3x^2+3y^2-2wx-4x^3, v = 6xy-2wy-4y^3\}. \quad (4.30)$$

Bifurcation set of the compactified elliptic umbilic The next step in order to understand how the equilibria change when moving in the catastrophe manifold is to study the bifurcation set, since it gives us information about the points in the control space for which sudden changes in the equilibria take place.

As in the elliptic umbilic catastrophe, the singularity set $\mathcal{S}$ is the set of points (x, y, u, v, w) in the catastrophe manifold such that the Hessian matrix is degenerate:

$$\frac{\partial^2 V_{u,v,w}(x,y)}{\partial x^2}\frac{\partial^2 V_{u,v,w}(x,y)}{\partial y^2} - \left(\frac{\partial^2 V_{u,v,w}(x,y)}{\partial xy}\right)^2 = 0,$$

i.e.

$$(6x + 2w + 12x^2)(-6x + 2w + 12y^2) - (-6y)^2 = 0.$$

Hence, $\mathcal{S}$ is given by the points in $\mathcal{M}$ that satisfy:

$$w^2 + w\left(6x^2 + 6y^2\right) + (36^2x^2y^2 - 18x^3 + 18xy^2 - 9x^2 - 9y^2) = 0. \quad (4.31)$$

Thus, the singularity set is the set in (x, y, u, v, w) space defined by:

$$\begin{aligned}\mathcal{S} = \quad & \left\{(x, y, u, v, w) \in \mathbb{R}^5 |\ u = -3x^2 + 3y^2 - 4x^3 - 2wx,\ v = 6xy - 4y^3 - 2wy,\right. \\ & \left. w^2 + w\left(6x^2 + 6y^2\right) + (36^2x^2y^2 - 18x^3 + 18xy^2 - 9x^2 - 9y^2) = 0\right\}.\end{aligned} \quad (4.32)$$

From equation 4.31, we can find two expressions for w as function of (x, y), which we will call $w_-(x, y)$ and $w_+(x, y)$:

$$w_-(x,y) = 3\left(-\left(x^2+y^2\right) - \sqrt{(x^4+y^4) + 2\left(-x^2y^2 + x^3 - xy^2\right) + x^2 + y^2}\right)$$

$$w_+(x,y) = 3\left(-\left(x^2+y^2\right) + \sqrt{(x^4+y^4) + 2\left(-x^2y^2 + x^3 - xy^2\right) + x^2 + y^2}\right)$$

Now, we can substitute these expressions in 4.32 and project into the control space to get that the bifurcation set is the union of the sets $\mathcal{B}_-$ and $\mathcal{B}_+$ where

$$\begin{aligned}\mathcal{B}_- = \quad & \{(u,v,w)\in\mathbb{R}^3| \\ & w = 3\left(-\left(x^2+y^2\right) - \sqrt{(x^4+y^4)+2\left(-x^2y^2+x^3-xy^2\right)+x^2+y^2}\right), \\ & u = -3x^2+3y^2-4x^3-2wx, \\ & v = 6xy-4y^3-2wy, \quad x,y\in\mathbb{R}\}\end{aligned} \tag{4.33}$$

$$\begin{aligned}\mathcal{B}_+ = \quad & \{(u,v,w)\in\mathbb{R}^3| \\ & w = 3\left(-\left(x^2+y^2\right) + \sqrt{(x^4+y^4)+2\left(-x^2y^2+x^3-xy^2\right)+x^2+y^2}\right), \\ & u = -3x^2+3y^2-4x^3-2wx, \\ & v = 6xy-4y^3-2wy, \quad x,y\in\mathbb{R}\}\end{aligned} \tag{4.34}$$

We have parametrised $\mathcal{B}$ by (x,y), i.e. given a value of (x,y), we can obtain a point $(u(x,y),v(x,y),w(x,y))$, i.e. $\mathcal{B}$ is a surface in the control space.

As we can see, the bifurcation set of the compactified elliptic umbilic is much more complicated than the bifurcation set of the elliptic umbilic, that we saw in Equation 4.28. We can, indeed, see how the bifurcation set changes when gradually changing the compactification term. With the purpose of seeing how the elliptic umbilic is modified by the compactification, we introduce a parameter λ in front of the last term of equation 4.29.

$$V_{u,v,w}(x,y) = x^3-3xy^2+w(x^2+y^2)+ux+vy+\lambda(x^4+y^4) \tag{4.35}$$

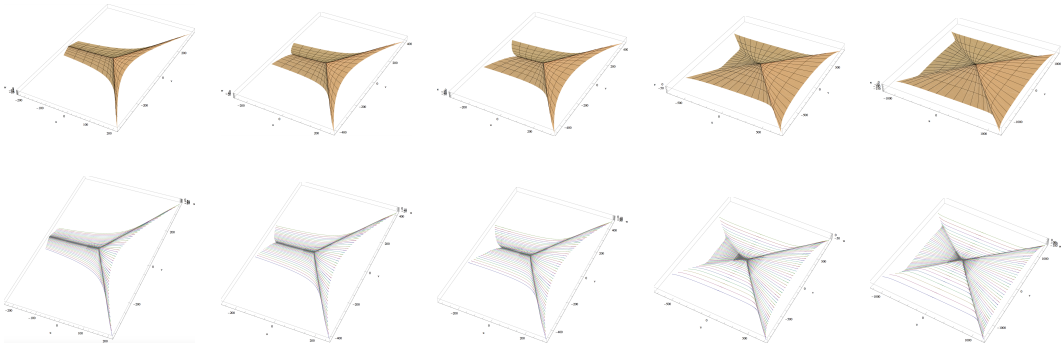
For this potential, the form of the bifurcation set would be $\mathcal{B}^\lambda = \mathcal{B}^\lambda_- \cup \mathcal{B}^\lambda_+$ where

$$\begin{aligned}\mathcal{B}^\lambda_- = \quad & \{(u,v,w)\in\mathbb{R}^3| \\ & w = 3\left(-\lambda\left(x^2+y^2\right) - \sqrt{\lambda(x^4+y^4)+2\lambda\left(-x^2y^2+x^3-xy^2\right)+x^2+y^2}\right), \\ & u = -3x^2+3y^2-4\lambda x^3-2wx, \\ & v = 6xy-4\lambda y^3-2wy, \quad x,y\in\mathbb{R}\}\end{aligned}$$

$$\begin{aligned}\mathcal{B}^\lambda_+ = \quad & \{(u,v,w)\in\mathbb{R}^3| \\ & w = 3\left(-\lambda\left(x^2+y^2\right) + \sqrt{\lambda(x^4+y^4)+2\lambda\left(-x^2y^2+x^3-xy^2\right)+x^2+y^2}\right), \\ & u = -3x^2+3y^2-4\lambda x^3-2wx, \\ & v = 6xy-4\lambda y^3-2wy, \quad x,y\in\mathbb{R}\}\end{aligned}$$

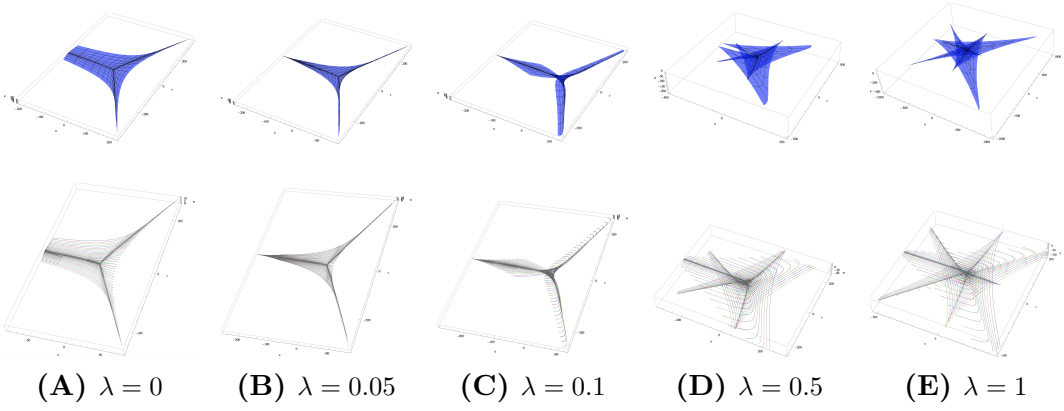
Figures 4.32 and 4.31 the sets $\mathcal{B}^\lambda_-$ and $\mathcal{B}^\lambda_+$ separately for different values of

λ. We can see how for $\lambda = 0$, the sets $\mathcal{B}^0_-$ and $\mathcal{B}^0_+$ correspond to the half cuspoidal cones as in Figure 4.28.



(A) $\lambda = 0$ **(B)** $\lambda = 0.05$ **(C)** $\lambda = 0.1$ **(D)** $\lambda = 0.5$ **(E)** $\lambda = 1$

Figure 4.31: *3D plots of the set $\mathcal{B}^\lambda_-$ in the (u, v, w) space for different values of the compactification parameter λ. For $\lambda = 0$, $\mathcal{B}^0_-$ is the half cuspoidal cone obtained for negative values of w*



(A) $\lambda = 0$ **(B)** $\lambda = 0.05$ **(C)** $\lambda = 0.1$ **(D)** $\lambda = 0.5$ **(E)** $\lambda = 1$

Figure 4.32: *3D plots of the set $\mathcal{B}^\lambda_+$ in the (u, v, w) space for different values of the compactification parameter λ. For $\lambda = 0$, $\mathcal{B}^0_+$ is the half cuspoidal cone obtained for positive values of w*

In order to explore the equilibria in different regions of the control space defined by the bifurcation set we will study the costant-w cross sections. The idea is to intersect $\mathcal{B}^\lambda$ with different planes $w = w_0$, so that we can study the bifurcations on the (u, v) plane. The bifurcation lines on the (u, v) plane will be given by the intersection of $\mathcal{B}_\lambda$ and the plane $w = w_0$. For example, if $\lambda = 0$ the intersection will be a three-cusped hypocycloid as in Figure 4.29. By changing the values of w_0 we can explore the entire bifurcation set.

The procedure to compute this intersection is the following. Given a value of $w = w_0$, it can be checked that the points (u, v, w) in the intersection of $\mathcal{B}_\lambda$ and

the plane $w = w_0$ can be parameterised as $(u(x, y, \lambda), v(x, y, \lambda), w_0)$ as we did before and they satisfy the equations:

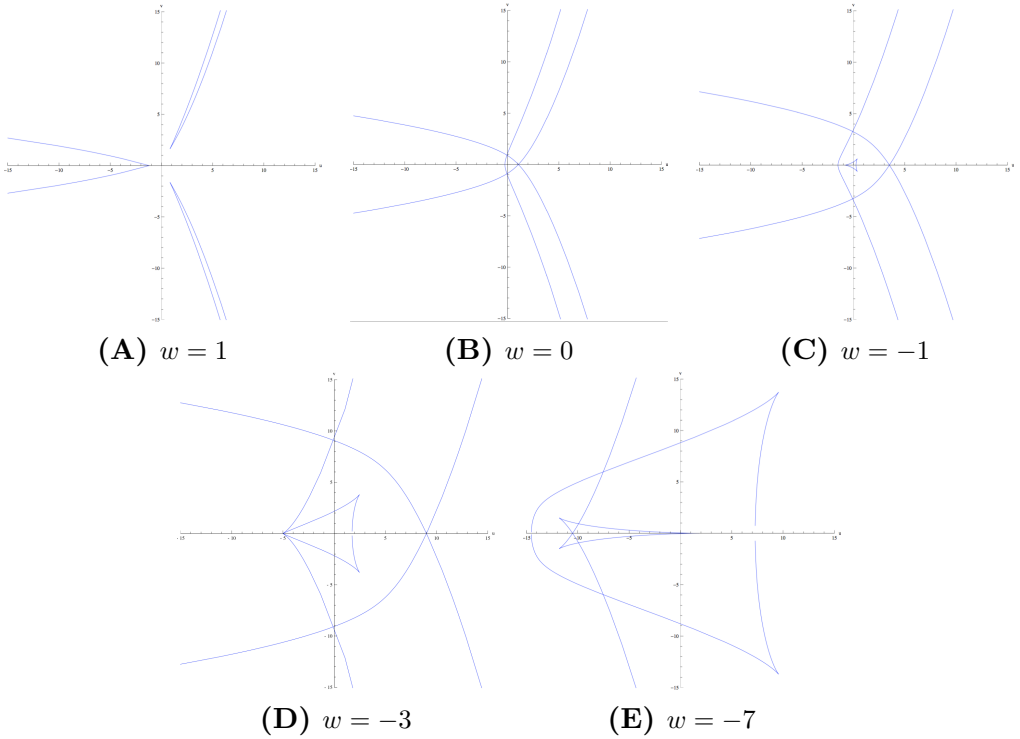
$$w_0^2 + w_0\left(6\lambda x^2 + 6\lambda y^2\right) + \left(36\lambda^2 x^2 y^2 - 18\lambda x^3 + 18\lambda x y^2 - 9x^2 - 9y^2\right) = 0, \quad (4.36)$$

$$u = -3x^2 + 3y^2 - 4x^3 - 2w_0 x,$$

$$v = 6xy - 4y^3 - 2w_0 y.$$

We can solve Equation 4.36 for y and write $y = y(x, w_0, \lambda)$. Therefore, fixed a value of λ and a plane $w = w_0$, the constant-w cross-section of $\mathcal{B}_\lambda$ is a one-dimensional set in the (u, v) space.

Figure 4.33 shows different constant-w cross sections of the bifurcation set $\mathcal{B}^\lambda$ on the (u, v) space, for different values of w and fixed $\lambda = 1$. For $w \in (0, -3)$ we can appreciate the three-cusped hypocycloid, which develops into a butterfly-type bifurcation for $w < -3$.



(A) $w = 1$ **(B)** $w = 0$ **(C)** $w = -1$

(D) $w = -3$ **(E)** $w = -7$

Figure 4.33: *Constant-w cross-sections of the bifurcation set $\mathcal{B}^\lambda$ of the compactified elliptic umbilic for different values of w and $\lambda = 1$.*

We have developed a Mathematica interactive application that allows the user to explore the bifurcation set by plotting the constant-w cross-sections of $\mathcal{B}^\lambda$

for different values of the parameters λ and w (see Figure 4.34). The user can specify the values of the parameters λ and w and the app will show this intersection on the (u, v) space. Moreover, the user can *click* on different regions of the (u, v) space and the app will also show the critical points and level-curves of the compactified elliptic umbilic for such values of the parameters.

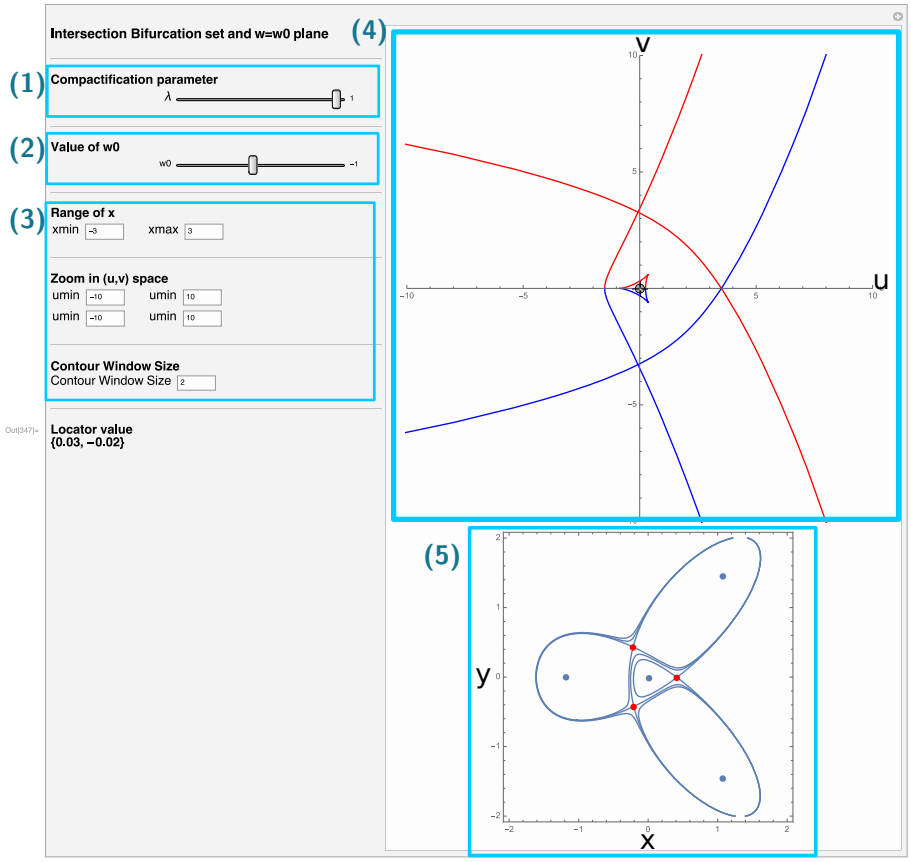


Figure 4.34: *Screenshot of the Mathematica interactive application designed to explore the bifurcation set of the compactified elliptic umbilic catastrophe. The app is divided into two halfs, on the left panels (1-3) the user can specify the options that will generate the plots on the right panels. (1) The user can specify the value of the compactification parameter λ. (2) The user can select the value of w that will be used to compute the constant-w cross-section of the bifurcation set $\mathcal{B}^{\lambda}$. (3) The ranges of the u, v, x, y values can be specified for the plots of the bifurcation set and the equilibria on the state space. (4) Plot of the constant-w cross-section of the bifurcation set $\mathcal{B}^{\lambda}$ on the (u, v) space, computed with the parameters specified by the user. The user can click on this plot to choose the (u, v) parameters for which the critical points and level curves will be plotted in (5). (5) Critical points and level curves curves of the potential function in Equation 4.35 where λ, w, u, v take the values specified by the user in (1),(2) and (4), respectively.*

Let us explore the bifurcations for $w = -1$, $\lambda = 1$. As shown in Figure 4.35, $\mathcal{B}^1_{-1}$ divides the (u, v) space into 8 connected components which are labelled from A to H. We will show that by changing the parameters u and v we can obtain the bifurcations that we think could help us to reproduce the process of vulval development.

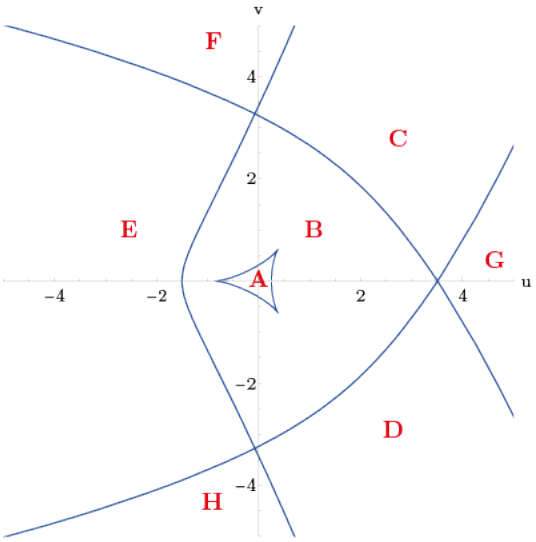


Figure 4.35: *Labelled regions in (u, v) space defined by the set $\mathcal{B}^1_{-1}$.*

Region A If the parameters u and v take values in region A, the potential corresponding to those parameter values has seven critical points: one maximum, three saddles and three minima. The change in the critical points and critical levels are shown in Figure 4.36.

Region B If the parameters u and v take values in region B, the potential corresponding to those parameter values has five critical points: three saddles and three minima (see Figure 4.37). Therefore, moving from Region A to region B produces a bifurcation of the maximum and a saddle.

Regions C, D and E If the parameters u and v take values in regions C, D or E the potential corresponding to those parameter values has two critical points: two minima and a saddle (see Figure 4.38). Therefore, moving from Region B to regions C, D or E produces a fold bifurcation between a minimum and a saddle.

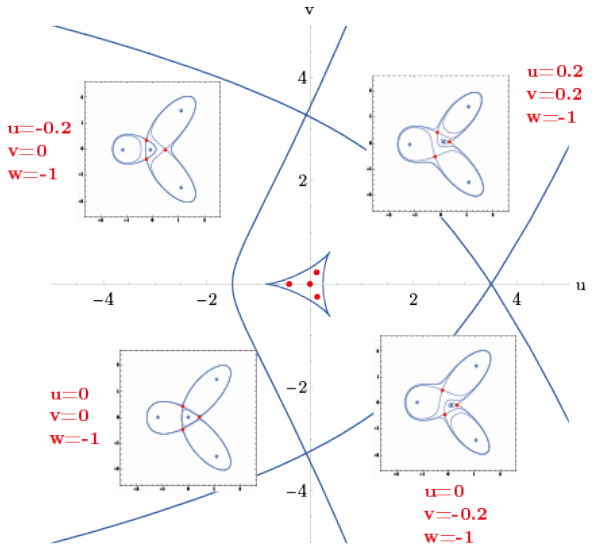


Figure 4.36: *Configurations of critical points of the compactified elliptic umbilic for $w = -1$, $\lambda = 1$, u, v taking values in region A. The red points in the (u, v) space correspond to the values of u and v for which we represent the critical points in the state space in the smaller plots. In the plots corresponding to the state space, the maximum and minimum points are coloured in blue, while the saddles are coloured in red.*

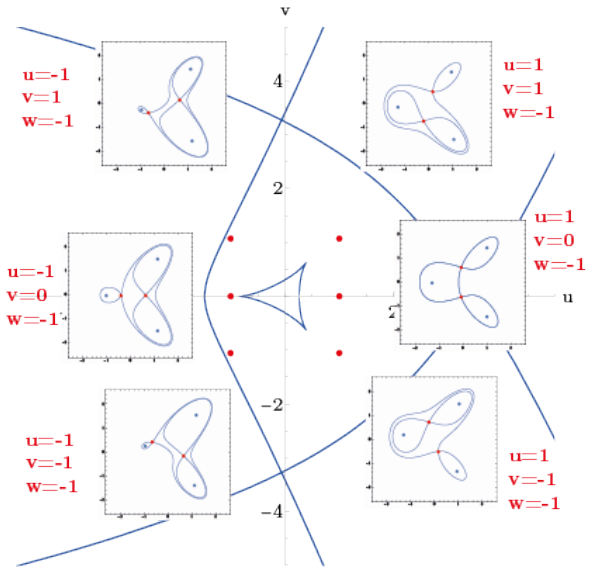


Figure 4.37: *Configurations of critical points of the compactified elliptic umbilic for $w = -1$, $\lambda = 1$, u, v taking values in region B. The red points in the (u, v) space correspond to the values of u and v for which we represent the critical points in the state space in the smaller plots. In the plots corresponding to the state space, the minimum points are coloured in blue, while the saddles are coloured in red.*

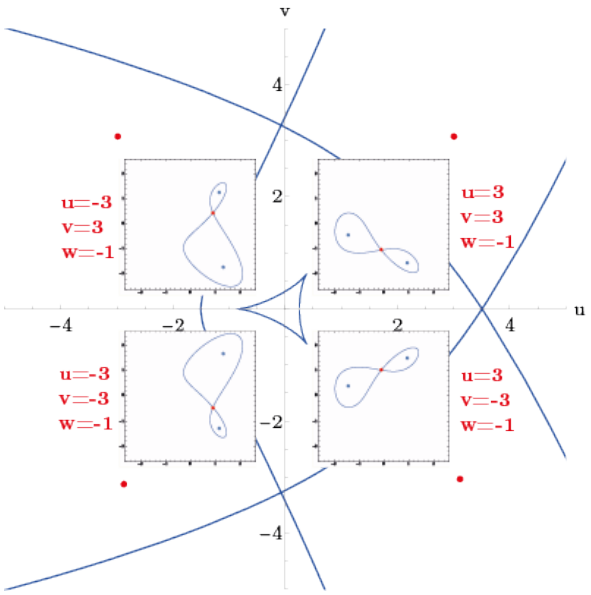


Figure 4.38: *Configurations of critical points of the compactified elliptic umbilic for $w = -1$, $\lambda = 1$, u, v taking values in regions C, D or E. The red points in the (u, v) space correspond to the values of u and v for which we represent the critical points in the state space in the smaller plots. In the plots corresponding to the state space, the minimum points are coloured in blue, while the saddles are coloured in red.*

Regions F, G and H If the parameters u and v take values in regions F, G and H the potential with to those parameter values has only one stable critical point. If u and v take values in region F, the critical point has positive x coordinate and negative y coordinate. If u and v take values in region G, the critical point has negative x. While if u and v take values in region H, the critical point has positive x coordinate and positive y coordinate.

This study shows that, for $w = -1$ and $\lambda = 1$, the compactified elliptic umbilic in Equation 4.35 is tristable, bistable or monostable depending on the values of u and v. Comparing the allowed bifurcations in the compactified elliptic umbilic and the topology we proposed in Chapter 3, the compactified elliptic umbilic has more flexibility and it resembles the topology chosen by Corson & Siggia (2012, 2017). Even though the compactified elliptic umbilic is not trifold symmetric as in Corson & Siggia (2012, 2017) it does have a symmetry, since $V_{u,v,w}(x, y) = V_{u,-v,w}(x, -y)$, and this could be helpful for assigning the fates. We conjecture that the compactified elliptic umbilic might be the potential system that can reproduce the process of vulval development if we assign fate 1° and 2° to any of the stable equilibria with positive x coordinate, and fate 3° to the minimum with negative x-coordinate.

In particular, in Figure 4.37, we can see that for parameter values $u =$

-1, $v = 0$, $w = -1$ the compactified elliptic umbilic contains the configuration of equilibria that we proposed for vulval development. From this configuration, by changing the parameter value of v the basins of attraction corresponding to 1° and 2° fates can be controlled. By changing the parameter u, the basin of attraction of the point that we think would correspond to 3° fate changes its relative size to the ones corresponding to 1° and 2° fates, which was the downside of the model that we proposed.

These are the reasons why we think that the compactified elliptic umbilic can be a suitable catastrophe for the modelling of vulval development. Further work would be needed to understand how to relate the control parameters to the biological signals and to find which experimental data sets could reproduce.

Chapter 5

A study of vertebrate trunk development

5.1 Biological background

There exist several vertebrate model organisms from which *Xenopus* (amphibian), zebrafish (fish), chick (bird) and mouse (mammals) stand out. However, in spite of their many differences, all vertebrates have a similar body plan. As the word *vertebrates* hints, they are characterised by the existence of a vertebral column in their bodies. The vertebral column defines the **antero-posterior axis** of the organism, the main body axis in vertebrates, starting at the head (anterior end of the axis), surrounding the spinal cord and, in many vertebrates, terminating at the posterior end in a post-anal tail. Although the morphogenetic processes establishing their body plans are different in early stages the development of structures such as the vertebral column and the central nervous system (CNS) is similar in all vertebrates (Wolpert et al. 2015). This means that understanding the process in one of the model organisms will shed light on the mechanism in the rest of vertebrates, such as humans. In this chapter, we will focus on the study of vertebrate trunk development in mouse.

In mouse, the generation of the antero-posterior body axis starts at E7.5 (7.5 days after the egg is fertilised) (Wolpert et al. 2015, Henrique et al. 2015) (see Figure 5.1). At this stage, the **anterior neural plate** (ANP), constituted by the forebrain, midbrain, hindbrain and some **anterior spinal cord** (aSC) as well as the **presomitic mesoderm** (PSM) are already part of the embryo. Cells coming from the posterior part of the embryo position themselves between the ANP and the posterior end of the embryo, pushing them apart and creating the antero-posterior axis. This way, the **spinal cord** (SC) and PSM (which develops into somites) are sequentially generated. One could think that all neural tissue (ANP and SC) would originate from a common progenitor that is distinct from the progenitors that give rise to mesodermal tissue. However, recent studies challenged this traditional idea.

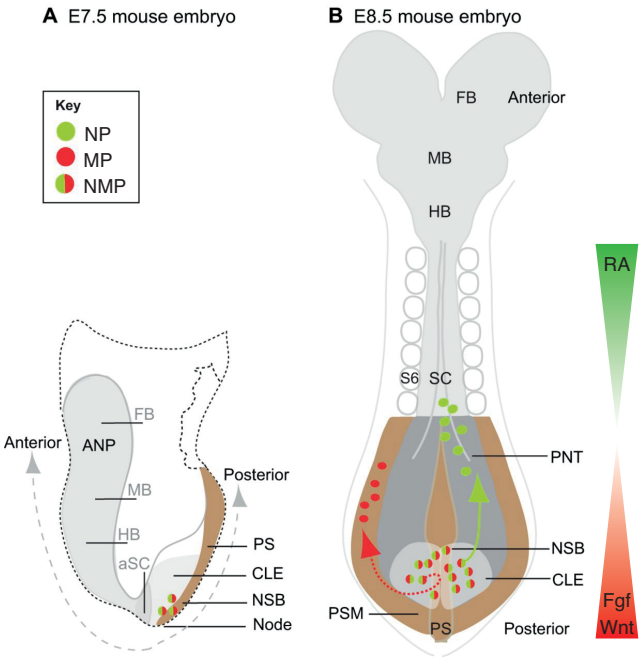


Figure 5.1: *Key features of the developing CNS, somitic mesoderm and neuramesodermal progenitors in the mouse embryo. (A) Sketch of E7.5 mouse embryo. At E7.5, the anterior neural plate (ANP) consists of the prospective forebrain (FB), midbrain (MB), hindbrain (HB) and some anterior spinal cord (aSC). Neuromesodermal progenitors (NMPs, coloured in red/green) are located in the node-streak border (NSB) in the anterior part of the primitive streak (PS, coloured in brown) and in the adjacent caudal lateral epiblast (CLE, coloured in light grey). (B) Sketch of E8.5 mouse embryo. The posterior gradients of Wnt and Fgf signals (in red), with source at the CLE, oppose the activity of retinoic acid (RA) produced by the developing somites. At E8.5, some NMPs have differentiated into new neural progenitors (NPs, coloured in green) which will contribute to the CLE and the preneural tube (PNT, coloured in dark grey), and into mesodermal progenitors (MPs, coloured in red), which contribute to the presomitic mesoderm (PSM, coloured in brown) which will then develop into somites. Adapted from Henrique et al. (2015).*

There is evidence that indicates that the SC and ANP originate from distinct progenitors residing in different regions of the developing embryo (Brown & Storey 2000, Tzouanacou et al. 2009, Gouti et al. 2015, Henrique et al. 2015). Tzouanacou et al. (2009) found that both spinal cord and paraxial mesodermal progenitors derive from the same population of bipotential **neuromesodermal progenitors** (NMPs), which are located in the caudal lateral epiblast (CLE) or stem zone adjacent to the node-streak border (NSB) (see Figure 5.1) (Brown & Storey 2000, Gouti et al. 2015, Henrique et al. 2015). NMPs are characterised by the co-expression of the early mesodermal-inducing transcription factor (TF) brachyury (T/Bra) and the epiblast and neural TF Sox2 (Gouti et al. 2015, Henrique et al. 2015). These studies suggest that spinal cord and paraxial mesoderm originate from a common progenitor that is

distinct from those which generate the anterior nervous system (Gouti et al. 2014). However, the molecular mechanisms that control the generation of mesodermal or posterior neural tissue, as well as the differences between the induction of posterior or anterior neural tissues, are not completely understood.

Several signals including members of the Fgf and Wnt families are known to be involved in the induction of mesodermal and neural tissue (Wilson et al. 2009, Streit et al. 2000, Wilson et al. 2000, Baker et al. 1999, Yamaguchi et al. 1999). In the mouse embryo, an anterior-posterior gradient of Fgf and Wnt signalling is created, with its high point at the CLE (Wilson et al. 2009, Gouti et al. 2015) (see Figure 5.1). These signals induce *Cdx* and *Bra* expression and CDX and BRA proteins in turn stimulate Wnt expression. This creates a feedback loop that sustains the NMP identity (Gouti et al. 2014, Henrique et al. 2015). Upregulation of CDX factors induces posterior Hox gene expression which establishes the posterior identity of NMP derived neural and mesodermal progenitors. These feedbacks need to be interrupted for a cell to exit the NMP state and differentiate into neural or mesodermal progenitors. Retinoic acid (RA), which is expressed in the somites, is involved in this process by inhibiting the expression of Fgf and Wnt (Wilson et al. 2009). As the NMP moves out of the CLE and approaches the somites, it is exposed to higher concentration of RA which could induce its differentiation. However, it is not clear whether increased RA signalling is sufficient to stop the NMP-sustaining feedback loops. The molecular details of the neural-mesoderm switch and how it is controlled *in vivo* remain unclear (Gouti et al. 2015).

In order to shed light on these mechanisms, biologists have studied these processes *in vitro*. In these studies, cultured stem cells derived from early stage embryos are exposed to different combinations of signals to study their changes in gene expression and investigate which types of cells they differentiate to. Recent studies have demonstrated that mouse and human embryonic stem cells (ESCs) as well as epiblast-derived stem cells (epiSCs) can be induced to NMP state *in vitro* (Henrique et al. 2015).

Gouti et al. (2014) and Turner et al. (2014) were the first investigators to succesfully provide evidence of the bipotential population of NMPs *in vitro*. Gouti et al. (2014) induced the mouse ESCs (mESCs) to an epiblast-like state by culturing them in the presence of FGF2 for two days and, following activation of Wnt signalling on the second day by adding CHIRON99021 to the culture, they obtain $Bra^+/Sox2^+$ NMP cells. Turner et al. (2014) achieved the same effect by culturing the mESC in CHIRON99021 and FGF2 from day 2 to day 3 of mouse ESC differentiation. The population of cells that they obtained co-expressed *Bra* and *Sox2* and could

generate neural and mesodermal cell types *in vitro*.

Neural precursor cells (NPCs) derived from NMPs downregulate *T/Bra* but maintain *Sox2* expression (Gouti et al. 2017). In contrast, mesodermal cells differentiating from NMPs downregulate the expression of *Sox2* but the expression of *Tbx6* is upregulated giving rise to **mesodermal progenitor cells** (MPCs) which, when commited to the paraxial presomitic mesodermal fate, also downregulate *Bra*.

Regarding the signals that induce NMP differentiation *in vitro*, Gouti et al. (2014) show that, when they differentiated epiblast-like cells without exposure to Wnt or with increased RA signalling, this generated neural precursors with anterior identity. By contrast, neural precursors derived from NMPs adopt more posterior fates. This suggests that Wnt signalling is necessary for posterior neural identity.

Understanding the GRN that balances NMP induction and differentiation *in vitro* could raise the possibility of taking advantage of NMPs for potential therapeutic use. Gouti et al. (2017) studied the differentiation of mESCs exposed to different signalling conditions using single-cell transcriptome analysis. They propose a GRN comprising *RA, Wnt, Sox2, T/Bra* and *Tbx6* as main variables in the network, which is obtained by fitting different GRN network topologies to the gene expression dynamics observed in their experiments. However, these dynamics that they fit are obtained in a qualitative way where the expression of a gene is described by being HIGH, MEDIUM or LOW.

In this chapter, we will propose new mathematical techniques that can help to quantitatively reverse engineer gene regulatory networks from gene expression analysis by applying these techniques to protein expression levels from flow cytometry data of mESCs differentiated into NMPs, neural and mesodermal cells *in vitro*.

First, we will explain the stem cell culture protocol and the experiments that our collaborators carried out. Secondly, we will introduce a novel visualization technique, called **Gene *i* ordered Gene *l* Expression** (GIGLE), that we have developed to understand how the expression of a given gene can affect the expression of other genes in a network. This visualisation technique will allow us to understand the data and to build a new GRN for the process of NMP induction and differentiation. Finally, we will propose some preliminary deterministic and stochastic mathematical models for this GRN, which can reproduce the data.

5.2 Biological data

5.2.1 Protocol

Mouse embryonic stem cells (mESC) are cultured in order to study their differentiation into neural progenitors or mesodermal cells, which is identified by the difference in expression of the genes *Tbx6*, *Sox1*, *Sox2*, *Cdx2* and *BRA*. mESC cells are cultured and maintained in an undifferentiated state on mouse embryo fibroblasts, and their development is then studied *in vitro* by differentiating them under defined culture conditions.

Experiments are performed for up to five days in which cells are cultured under different signalling conditions. The protocol for each day goes as follows:

Day 0 mouse ES cells are plated in basal medium containing 10ng/ml bFGF and LGK974 (inhibitor of WNT secretion).

Day 2 mESCs are in **epiblast** state. Wnt signalling was activated with CHIRON99021 (CHIR) or not.

Day 3 Fgf was removed from the media. Depending on the experiment, either CHIR was removed or sustained.

Day 3 to 5 Same experimental conditions were maintained.

A proportion of cells in the culture is fixed for immunofluorescence staining every number of hours from day 2, while the rest of cells continue proliferating under the chosen experimental regime. The fixed cells are stained with fluorescent dyes conjugated to specific antibodies that target each protein of interest (TBX6, SOX1, SOX2, CDX2 and BRA) for the acqusition of the corresponding fluorescence level with a flow cytometer. The amount of fluorescence, ideally, is directly proportional to the corresponding level of protein expression in the cell.

It is important to note that this technique does not provide data about the evolution of the protein expression of a particular cell in time since the fixed cells are different at each time point. Instead, it gives data about the overall dynamics of protein expression within populations of cells that have developed under the same experimental regimes.

There are some caveats in the staining that are worth mentioning. Some antibodies seem to *perform* better than others, in the sense that for some poorly performing antibodies the raw fluorescence measurement does not reliably represent the corresponding amount of protein in the cell. For example, we have noticed that

the signal of certain antibodies is inversely proportional to the number of fixed cells, lower number of cells producing higher background non-specific signal. However, prior knowledge obtained from complementary techniques not subject to these technical errors can be used to deal with these problems. In Subsection 5.3.1 we will propose a method to process the data that accounts for these errors.

5.2.2 Data sets

Following the experimental procedure explained in 5.2.1, our collaborators provided us with three data sets that contain information about the protein expression of populations of mESC at several times of their development under different experimental conditions.

Data set 1 This data set contains eleven samples with the fluorescence levels of three WT mESCs cultures exposed to three different experimental conditions. Samples are taken at days 2, 2.5, 3, 3.5 and 4 in the three cultures. Each sample is an $N_1 \times 5$ matrix that contains numerical information about the fluorescence levels of five proteins TBX6, SOX1, SOX2, CDX2 and BRA for $N_1 = 40,000$ cells. The sample taken at day 2 is identical for all the cultures since it is taken before CHIR is or is not added to the media. The first culture is never treated with CHIR in any of the days (No CHIR condition). This provides five samples (one for each day). The second culture is treated with CHIR from day 2 to day 3, after which CHIR is removed from the culture (CHIR 2-3 condition). This provides four samples (from which two samples correspond to cells proliferating in CHIR at days 2.5 and 3, and two samples correspond to days 3.5 and 4 of cells which proliferated in CHIR from day 2 to day 3 and without CHIR from day 3 onwards). The third culture is exposed to CHIR from day 2 to day 4 (CHIR condition). This provides two new samples which correspond to days 3.5 and 4 of cells which proliferated in CHIR from day 2 to day 4. This gives eleven samples in total from which we can obtain data about the changes in protein expression of mESCs proliferating under three experimental conditions: mESCs proliferating with no CHIR for 5 days (No CHIR condition), mESCs proliferating with CHIR from day 2 to day 3 and without CHIR from day 3 onwards (CHIR 2-3 condition) and mESCs proliferating with CHIR for 5 days (CHIR condition).

Data set 2 This data set contains four samples with the fluorescence levels of populations of WT mESCs exposed to CHIR from day 2 to day 5, i.e. under CHIR condition. The samples are taken at days 2.5, 3, 4 and 5 and each sample is

an $N_2 \times 4$ matrix that contains numerical information about the fluorescence levels of four proteins TBX6, SOX2, CDX2 and BRA for $N_2 = 50,000$ cells taken at the corresponding days.

Data set 3 This data set contains sixteen samples with the fluorescence levels of populations of $Bra^{-/-}$ and WT mESCs cells that are exposed to No CHIR, CHIR 2-3 and CHIR experimental conditions from day 2 to day 3.3. The samples are taken at days 2 (two samples), 2.3 (four samples) , 2.7 (four samples) and 3.3 (six samples). Both $Bra^{-/-}$ and WT mESCs cells are exposed to No CHIR, CHIR 2-3 and CHIR experimental conditions as explained in Data Set 1. Each sample is an $N_3 \times 5$ matrix that contains numerical information about the fluorescence levels of five proteins TBX6, SOX1, SOX2, CDX2 and BRA for $N_3 = 5,000$ cells taken at the corresponding days.

The goal is to obtain information about the gene regulatory network that controls neural and mesoderm specification from the listed data sets. In the following sections we will explain how we analyse them and extract the key knowledge that will allow us to construct a possible GRN.

5.3 Data Analysis

5.3.1 Data Processing

When visualising the samples explained in Subsection 5.2.2, we observed that the distributions of fluorescence levels for the proteins were normally bimodal or with long tails as in Figure 5.2.

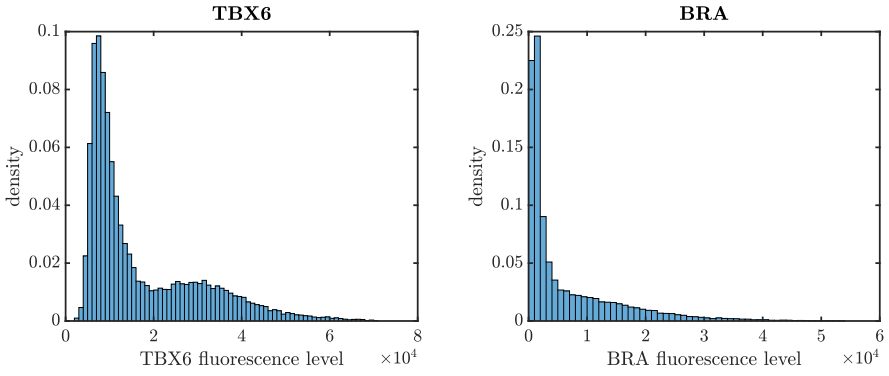


Figure 5.2: *Distributions of fluorescence levels of TBX6 and BRA obtained under CHIR condition in Data Set 1, at days 4 and 3 respectively.*

The peaks of the distributions normally correspond to different cell types. It is a common problem in flow cytometry data analysis that populations of cells of the same type (same gene expression) are assigned different fluorescence levels when measured in different rounds with the flow cytometer. This is due to the poor signal of the antibodies (as explained in Subsection 5.2.1), different settings in the machine or other uncontrollable or simply unknown factors. This means that peaks corresponding to the same type of cells can appear shifted from sample to sample, making the analysis difficult.

In our case we observe that the lower peaks of the distributions of fluorescence values of TBX6, CDX2 and BRA are located at different fluorescence values in different samples (see Figure 5.3). We know from single-cell qPCR that the low peaks of these genes correspond to a negative population of cells that do not express the corresponding gene and therefore cannot express the corresponding protein, i.e. even though the flow cytometer assigns a non zero fluorescence level to these cells their fluorescence level is indeed zero.

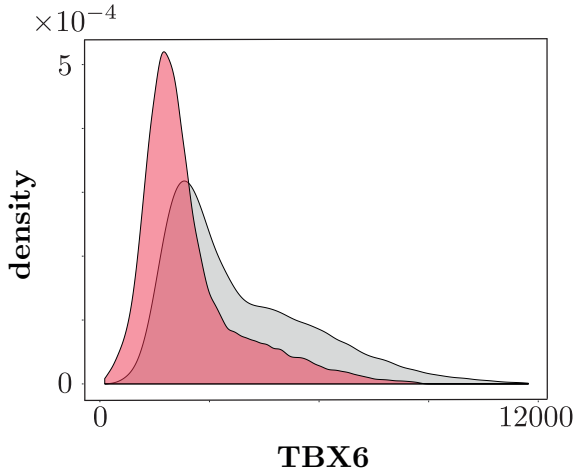


Figure 5.3: *Distribution of fluorescence levels obtained for TBX6 under CHIR condition in Data Set 1, at days 4 (red) and 5 (grey) respectively. The lower peaks corresponding to the negative population of cells that do not express TBX6 appear shifted. We can also observe that the support of the distributions look shifted as well.*

In order to solve this problem, our collaborator Dr. Robert Blassberg developed a method that scales the data so that the peaks corresponding to the known negative populations are aligned, making the comparison between samples possible. The method is applied to each data set separately, and it proceeds as follows:

1. Subsample the data that is within 3 standard deviations around the mean. This step trims the data to remove outliers.

2. Downsample the data set so that it contains the data of N cells at each day, where N is the number of cells of the smallest sample in the data set.

3. Choose a protein P for which the negative populations need to be aligned.

4. Choose a sample in the data set whose distribution of fluorescence levels for protein P is clearly bimodal. Let us say that this is the i-th sample in the data set. This will be the sample that will be used to scale the rest of the samples in the data set.

5. Fit a mixture of two gaussian distributions to the sample, i.e. decompose such distribution as the sum of two gaussian distributions. Compute the mean of the two gaussian distributions and denote by μ_i the smallest one of them. This is the mean of the negative population (see Figure 5.4).

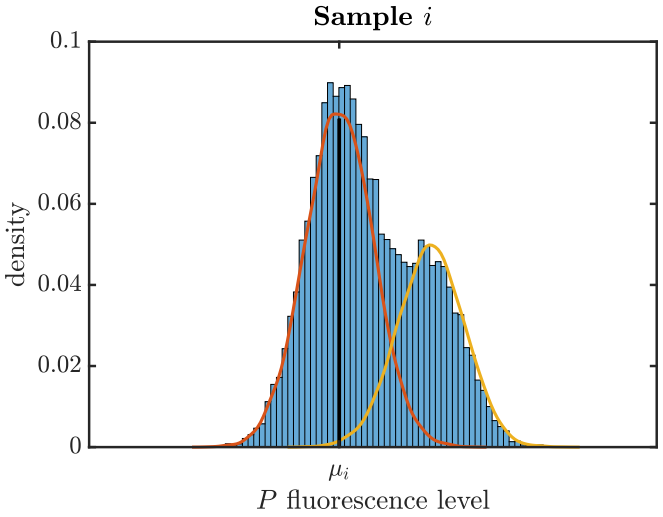


Figure 5.4: *Decomposition of a bimodal distribution into a mixture of two gaussian distributions to find the mean of the negative population. The histogram is obtained from the P fluorescence levels in sample i. The mixture distributions are represented by orange and yellow lines. The orange distribution is the one with the lowest mean, μ_i, and therefore is considered the distribution of the negative population.*

6. For each one of the other samples do the following:

   (a) Decompose the distribution of protein P levels in the new sample j as the sum of two gaussian distributions. Compute the mean of the two gaussian distributions and denote by μ_j the smallest of the two means. This is the mean of the negative population in sample j.

   (b) Compute the ratio μ_i/μ_j. This will be the scaling factor.

(c) Multiply the P fluorescence levels in sample j by the scaling factor μ_i/μ_j. This way the j-th sample is scaled so that the mean of the negative population is now μ_i, therefore it is aligned to the negative population of sample i.

Figure 5.5 shows the result of applying this method to the TBX6 data that were plotted in Figure 5.3. We use this method to rescale TBX6, CDX2 and BRA levels whithin each of the data sets provided.

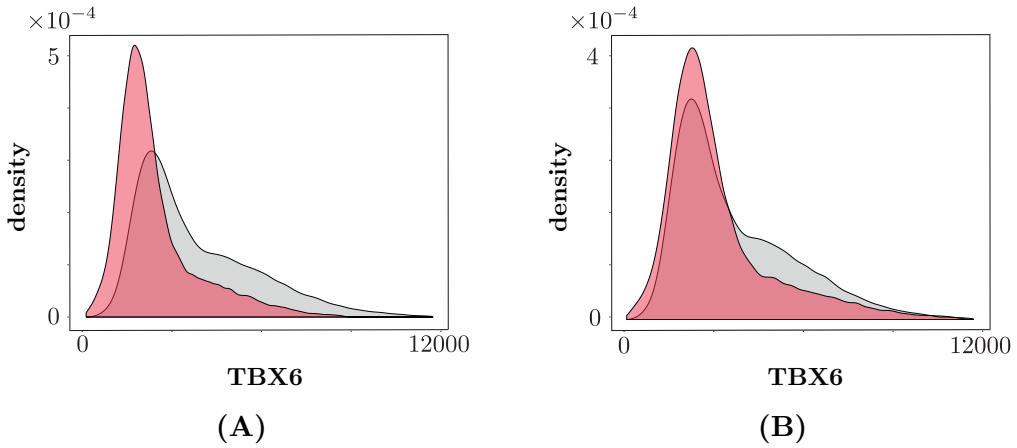


Figure 5.5: *Result of applying the rescaling method to the TBX6 fluorescence levels in samples CHIR day 4 (red distribution) and CHIR day 5 (grey distribution) in Data Set 1. (A) Distributions of the unscaled levels. We can see that the negative populations are not aligned. The distribution in grey is chosen as the sample i of reference. (B) Distributions of the rescaled levels. The red distribution has been rescaled to match the negative population of the grey distribution. We can see that the negative populations are now aligned, and also the support of the distributions are the same.*

It is important to note that the ranges of fluorescence levels are different for each data set. We tried to normalise the data sets to [0,1] by summing the fluorescence levels of each protein by its minimum fluorescence level and dividing by its maximum fluorescence level in the data set. However, since the minimum and maximum values are highly stochastic, this normalisation did not give good results. We decided then to work with unnormalised data sets, which gave good preliminary results as we will explain in the following sections.

5.3.2 Data Visualisation with GIGLE

By taking advantage of the provided experimental data, our goal is to gain knowledge about the gene regulatory network controlling the process.

The data sets listed in Subsection 5.2.2 are very high dimensional since they

contain fluorescence levels of at least four proteins for tens of thousands of cells in several days under different conditions, so visualising these data is not trivial. Different methods have been developed for visualising and analysing flow cytometry data. These can be grouped into dimensionality-reduction methods, such as principal component analysis (PCA) or t-stochastic neighbour embedding (t-SNE), and clustering methods, such as SPADE (spanning tree progression of density normalized events) algorithm or FlowSOM (flow cytometry data analysis using self-organizing maps) (Saeys et al. 2016). Dimensionality-reduction methods aim to find a lower-dimensional description of the (usually) high-dimensional data. For example, t-SNE preserves similarities between the data points when projecting it onto the lower-dimensional space, so that cells that are similar in the original space are similar in the low-dimensional space (although the opposite might not be true) (Van Der Maaten & Hinton 2008). On the other hand, clustering-methods aim to group cells into clusters in the high-dimensional space that could correspond to different cell types in order to, then, use visualization algorithms to represent these clusters in a lower-dimensional space. SPADE, as an example of such methods, applies hierarchical clustering to group the cells into cell-type clusters that are then represented in a spanning tree where similar clusters are closer in the tree (Qiu et al. 2011).

Here we present a new visualization technique, **Gene *i* ordered Gene *l* Expression** (GIGLE). GIGLE allows to observe how the genes interact with each other in the network and how different types of cells can arise from these interactions. It is particularly useful for exploring the dynamics of the genes by looking at the evolution of the data in time. It can be applied to gene or protein expression data, but on the following subsections we will provide a description of the method by applying it to the flow cytometry data sets mentioned in Section 5.2. As a result of this analysis, we will be able to propose a gene regulatory network for the neuromesodermal differentiation of mESCs.

5.3.2.1 Description of GIGLE as applied to the experimental data

The GIGLE visualisation technique provides a way of comparing the fluorescence levels of two proteins in the population of cells, which will give information about how these two genes interact in the GRN.

Suppose that we were interested in understanding how the fluorescence levels of the i-th protein and the levels of the l-th protein are related in a certain sample. Similarly, it could be applied to understand how the gene expression values of an i-th gene and the gene expression values of an l-th gene were related in a sample. We can visualise this relationship with GIGLE, which comprises two steps: an ordering

step and a binning step.

Ordering For each one of the N cells in the sample, a value for protein i and a value for protein l are avaliable. The idea is to order the cells by their i-th protein fluorescence level in an increasing manner. We use this to index the cells from 1 to N with 1 having the lowest intensity. Then, for each cell the fluorescence level of protein l is plotted against the index of the cells. In turn, this orders the fluorescence levels of protein l by the levels of protein i. The data, when ordered this way, gives information about how the level of protein l changes as the level of protein i increases (see Figure 5.6).

Binning Normally, the number of cells N in the sample is very high so plotting the ordered data does not give too much information. The next step, once the data is ordered by protein i, is to bin the data. We define bins B_j, of size M, where bin B_j contains the cells whose indexes are between $(j-1)M$ and jM. Each bin is associated with an intensity given by the mean protein l intensity of the cells in the bin. We plot this bin intensity against the bin index j. One can also plot error bars with the 95% confidence interval of the means as in Figure 5.6.

GIGLE plots allow the observation of the correlation between two protein levels. In the case in which there is a repression between gene i and gene l, we will observe that the binned values of protein l are low in bins where the value of protein i is high and the other way around (as in Figure 5.6, where we can see that for SOX2 low cells, TBX6 expression is higher).

Note that ordering step must be done before binning so that only similar cells are averaged. If the binning is done before ordering, very different cells contribute to the value of the same bin and, therefore, the information is lost.

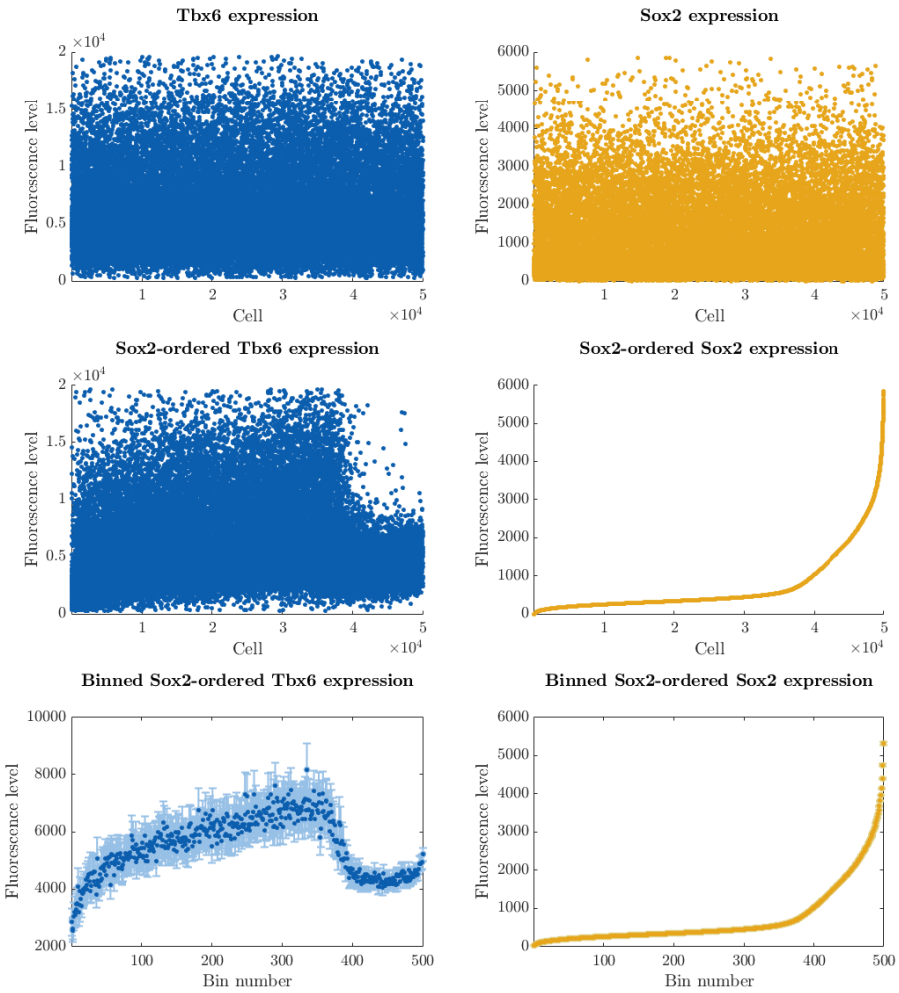


Figure 5.6: *Steps to obtain a GIGLE plot. On the first row, the scatter plots of the TBX6 data (in blue) and SOX2 data (in yellow). On the x-axis is the original order of the cells in the sample corresponding to day 5 under CHIR condition in Data Set 2. On the second row of plots, the scatter plots of the SOX2-ordered TBX6 data (in blue) and SOX2-ordered SOX2 data (in yellow) against the cell index. Each point in the scatter plot corresponds to one cell in the sample. On the last row of plots, the binned SOX2-ordered TBX6 data (in blue) and binned SOX2-ordered SOX2 data (in yellow). Each point in the scatter plot corresponds to the average expression in a bin of size 100. The 95% confidence intervals of the data in each bin are represented with error bars in a lighter colour.*

GIGLE methodology can be used for plotting the data in more than one dimension, using the fluorescence levels of each protein in each bin as coordinates. If the sample contains data of more than 4 proteins, up to four-dimensional (4D) GIGLE plots can be obtained (see Figures 5.7–5.9)

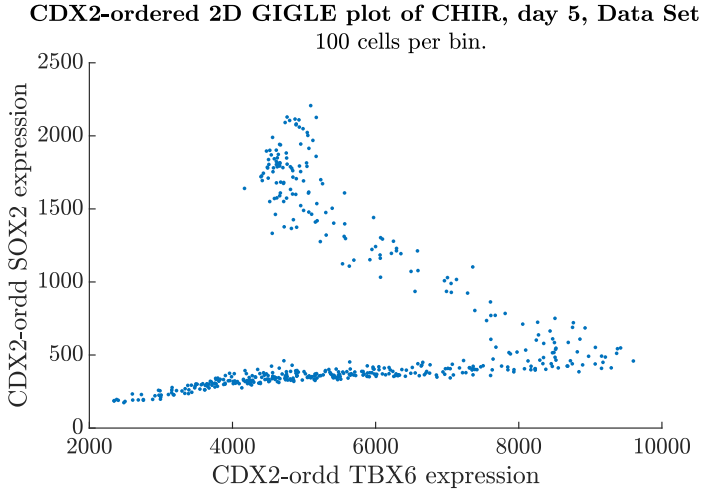


Figure 5.7: *Example of two-dimensional GIGLE plot obtained from ordering TBX6 and SOX2 expression by CDX2 expression in the sample corresponding to day 5 under CHIR condition in Data Set 2. It might surprise the reader that it looks like a bifurcation set where the unstable brach is visible. This has to do with the stochastic nature of the data and we will discuss more about it in Section 5.7.*

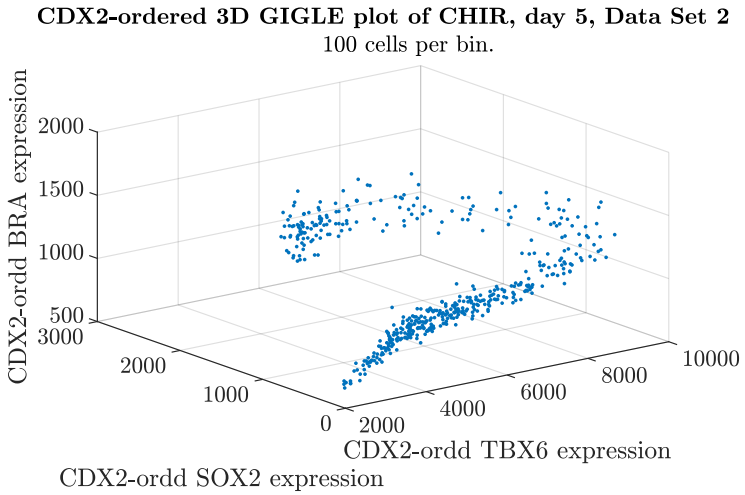


Figure 5.8: *Example of three-dimensional GIGLE plot obtained from ordering TBX6, SOX2 and BRA expression by CDX2 expression in the sample corresponding to day 5 under CHIR condition in Data Set 2.*

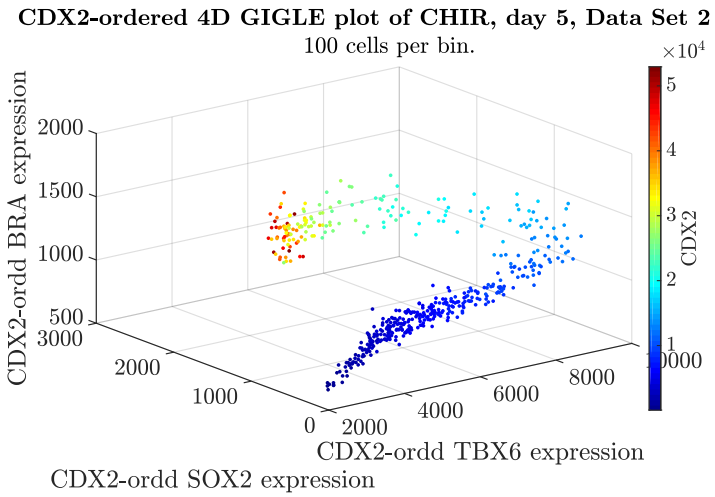


Figure 5.9: *Example of four-dimensional GIGLE plot obtained from ordering TBX6, SOX2, CDX2 and BRA expression by CDX2 expression in the sample corresponding to day 5 under CHIR condition in Data Set 2. The points are coloured by their relative CDX2 expression. We can see that there are two types of cells, one with SOX2 and CDX2 high and another one with SOX2 and CDX2 low.*

When visualizing the data using GIGLE plots, it is important to repeat the plots with different bin sizes (see Figure 5.10). A very big bin size could mask some information. For example, it could happen that for a range of values of protein i two populations of cells coexist, one with high protein l expression and other with low protein l expression. If the bin size is too high, the GIGLE plot will show only one population of *middle* protein l expression, that is obtained by binning cells with high and low protein l expression in the same bin. In spite of this, the result will always show the trend of the data with respect to protein i because this *middle* value of protein l expression will reflect what the proportion of cells in each one of the two populations is. For example, in the sorted but not binned GIGLE plots of Figure 5.6 (plots on the second row), we can see that for low values of SOX2, there exist cells with high and low TBX6 expression, however for high values of SOX2, only cells with low TBX6 exist. When that data are binned (plots on the third row of Figure 5.6), the GIGLE plot shows a higher average of TBX6 expression for values of SOX2 low than in the region when SOX2 is high. The GIGLE plot must therefore be interpreted as a summary of the data.

GIGLE plots are especially useful when data are available for diferent time points, as one can obtain information about the dynamics of the gene expression, as we will explain in the following subsections.

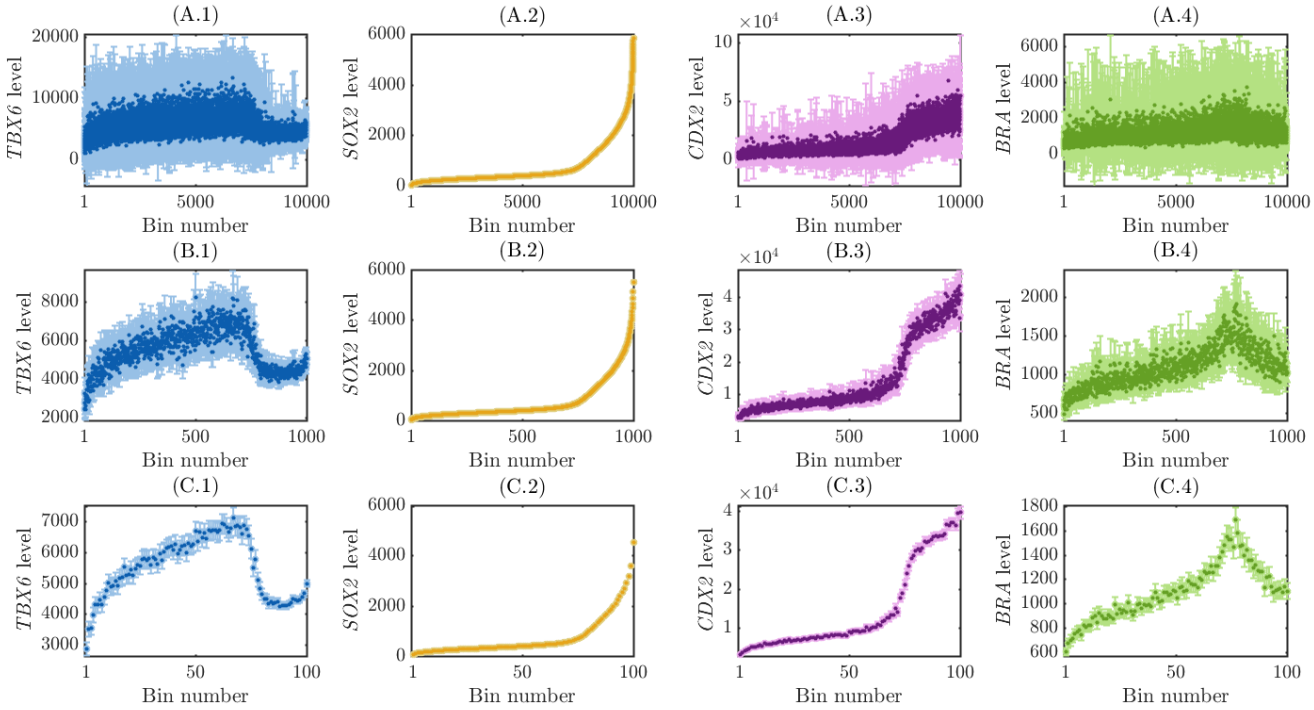


Figure 5.10: *Comparison of the SOX2-ordered 1D GIGLE plots for TBX6, SOX2, CDX2 and BRA expressions in the sample corresponding to day 5 under CHIR condition in Data Set 2 with different bin sizes. Row (A) corresponds to 5 cells per bin, i.e. 10,000 bins. Row (B) corresponds to 50 cells per bin, i.e. 1,000 bins. Row (C) corresponds to 500 cells per bin, i.e. 100 bins. In each GIGLE plot, the means of the data in each bin is represented by a dot of dark colour blue (yellow, purple, green) for TBX6 (SOX2, CDX2, BRA respectively).The 95% confidence interval of each bin is represented with an error bars in the corresponding light colour.*

5.3.2.2 4D GIGLE plots as a Matlab GUI

In order to facilitate the data visualisation, we develop a Matlab GUI that allows the user to explore the data sets described in Subsection 5.2.2 using GIGLE plots (see Figure 5.11).

With this GUI, the user can choose the data set, the experimental condition and the day of the sample that will be plotted. The user needs to choose the protein i by which the data will be ordered, and the number of bins that will be used to produce the GIGLE plot. Then, the data are plotted in a three-dimensional scatter plot in which the user can also select the proteins that will be represented in each axis. The bins will be coloured according to their relative value of the expression of the protein i. The limits of the axis can be chosen to be adapted to the sample that is being plotted, to be fixed to the range of values obtained under a certain experimental condition or to be fixed to the range of values in the entire data set. For an easier comparison of the evolution of the data across the days, the GUI plots all the samples corresponding to the same experimental condition in a data set together, giving opaque colour scheme to the sample chosen by the user and shadowing the rest of samples.

This tool helps to explore how the gene expression of the cells changes in time and observe the differentiation of the mESCs into NMP, mesodermal progenitors and neural progenitors, as we will explain in the following subsection.

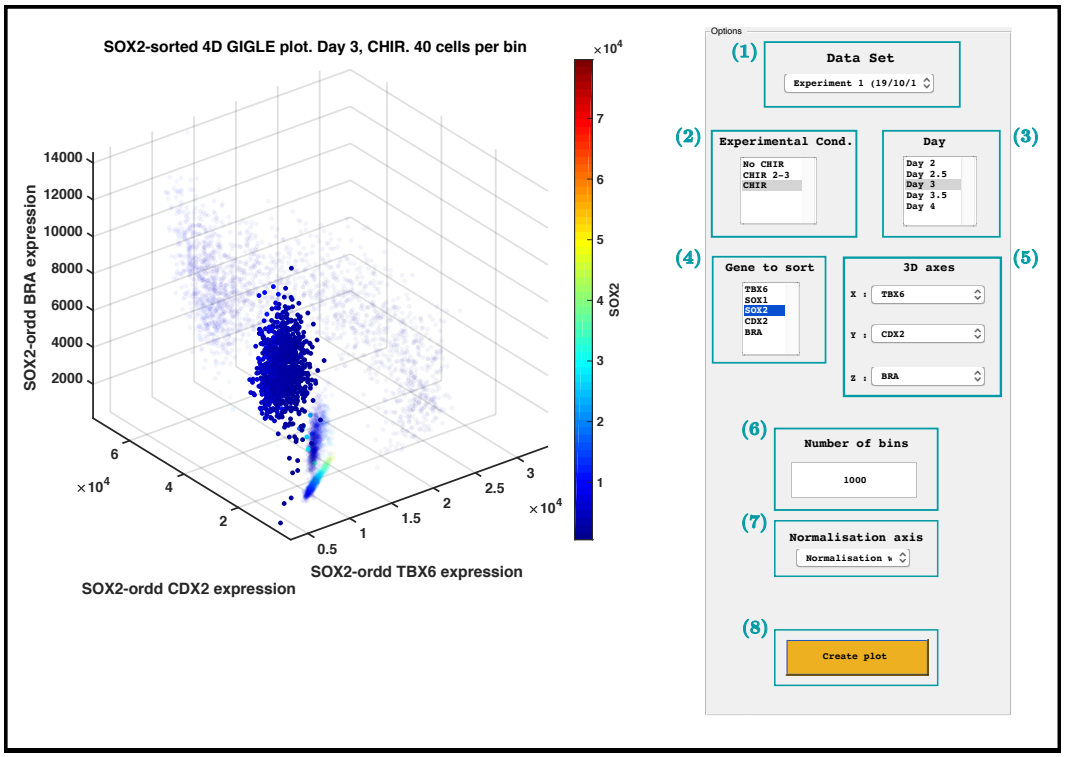


Figure 5.11: *Screenshot of the GUI for the visualisation of GIGLE plots of the data sets available. The window is divided into two parts: the one on the right contains a panel with the options to create a 4D GIGLE plot, and the one on the left contains the corresponding GIGLE plot obtained from the options chosen. (1) The user can choose the data set from which the sample will be obtained. The values in the boxes (2)-(5) adapt to the data set chosen. For example, in some data sets only CHIR condition is available. (2) Experimental condition in which the data that will be plotted was obtained. (3) Day of the sample. (4) Protein i by which the data will be sorted to produce the 4D GIGLE plot. The points in the scatter plot will be coloured according to the value of the chosen protein i. For example, in this case the data is SOX2-ordered and the bins are coloured according to their mean SOX2 value. (5) Genes that will be represented on each axis. For example, in this GIGLE plot, SOX2-ordered TBX6 (CDX2, BRA) expression is represented on the x-axis (y-axis, z-axis respectively). (6) Number of bins that will be used in the GIGLE plot. (7) Limits of the axes. In order to compare data from different days or under different conditions the GUI allows the user to choose how the limits of the axes are computed. For example, the limits of the axes can be fixed to be the maximum and minimum values of the corresponding protein in the whole data set. This allows the user to easily compare the change of the expression of a gene under different experimental conditions or different days. (8) Button that the user needs to press in order for the GUI to create the GIGLE plot corresponding to the chosen options.*

5.3.3 Insights about mESC differentiation under different Wnt regimes with GIGLE

We study the evolution of the gene expression of *Tbx6*, *Sox2*, *Cdx2* and *Bra* in WT epiblast mESCs under No CHIR, CHIR 2-3 and CHIR conditions from day 2 to day 4. For this purpose, we will concentrate on Data Set 1 since it is the only data set that contains data for the three conditions for all days. This allows us to observe the relative difference of protein expression under the three conditions, circumventing the data normalisation problem. However, we will refer to the samples from the other data sets in some cases. We also decide, on a first approach, not to study *Sox1* expression since it is a neural marker such as *Sox2*. We will get back to this in Section 5.7.

Figures 5.12, 5.13 and 5.14 show the 1D GIGLE plots for the data obtained under No CHIR, CHIR 2-3 and CHIR conditions, respectively. We order the data by SOX2 protein levels since we find that *Sox2* is the only gene that has strong interactions with all the other genes. The axes have been fixed for the entire data set so that the relative differences of protein expression under the different conditions can easily be noticed. Figures Figures 5.15–5.23 show the 1D GIGLE plots of different data sets that will be useful to show gene interactions. We will refer to them in this and the following sections.

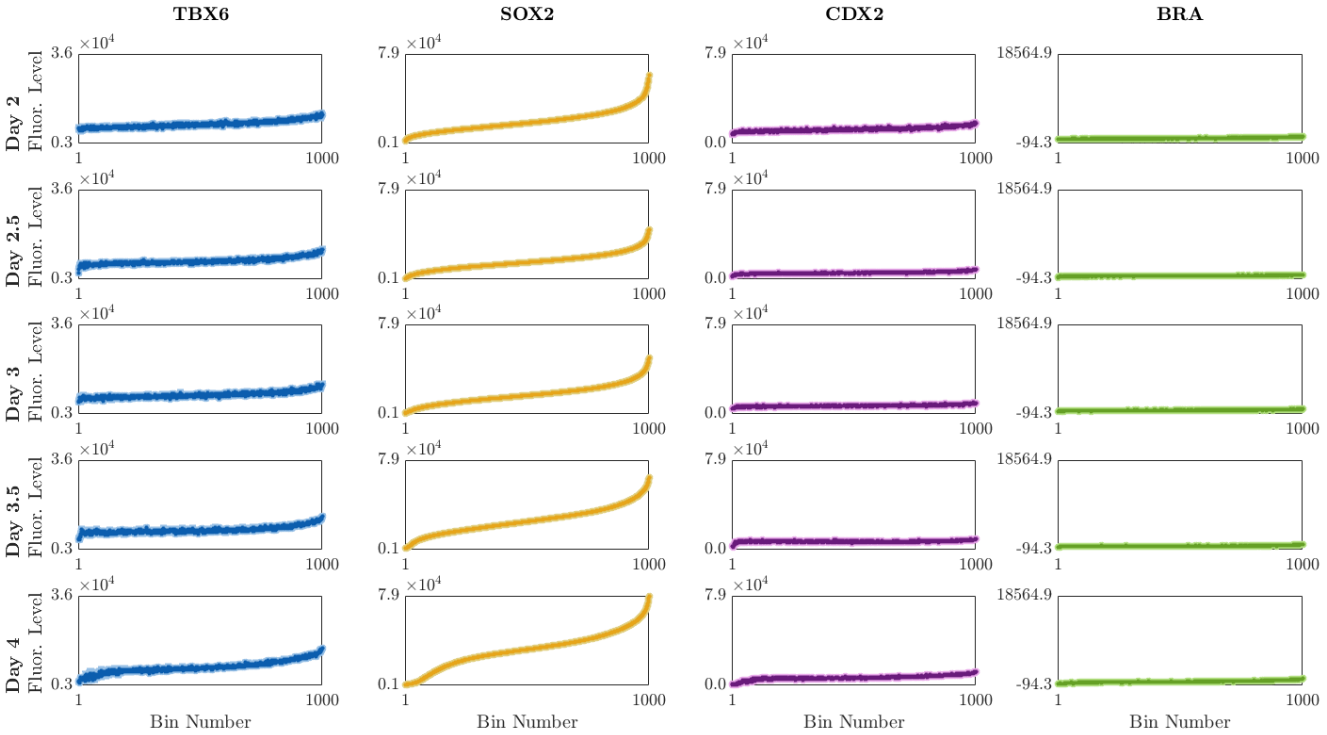


Figure 5.12: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under No CHIR condition. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

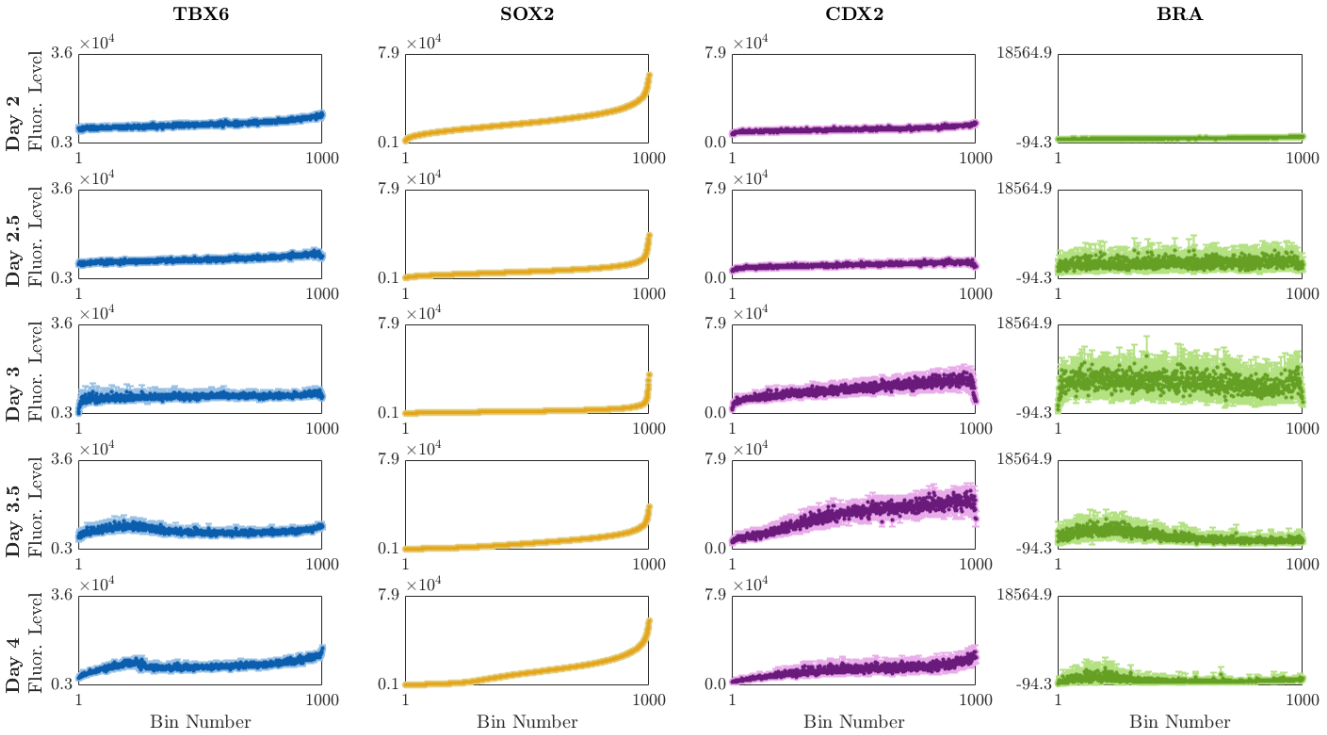


Figure 5.13: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under CHIR 2-3 condition. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

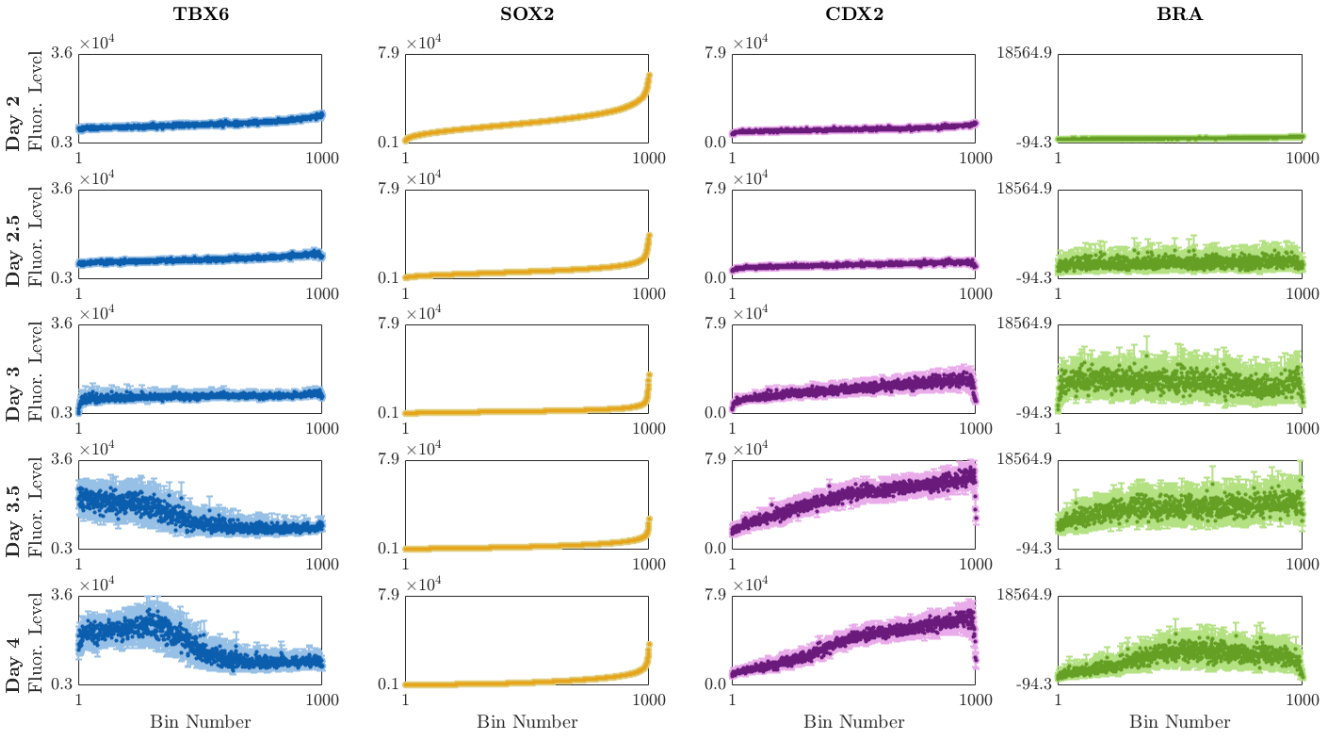


Figure 5.14: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under CHIR condition. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

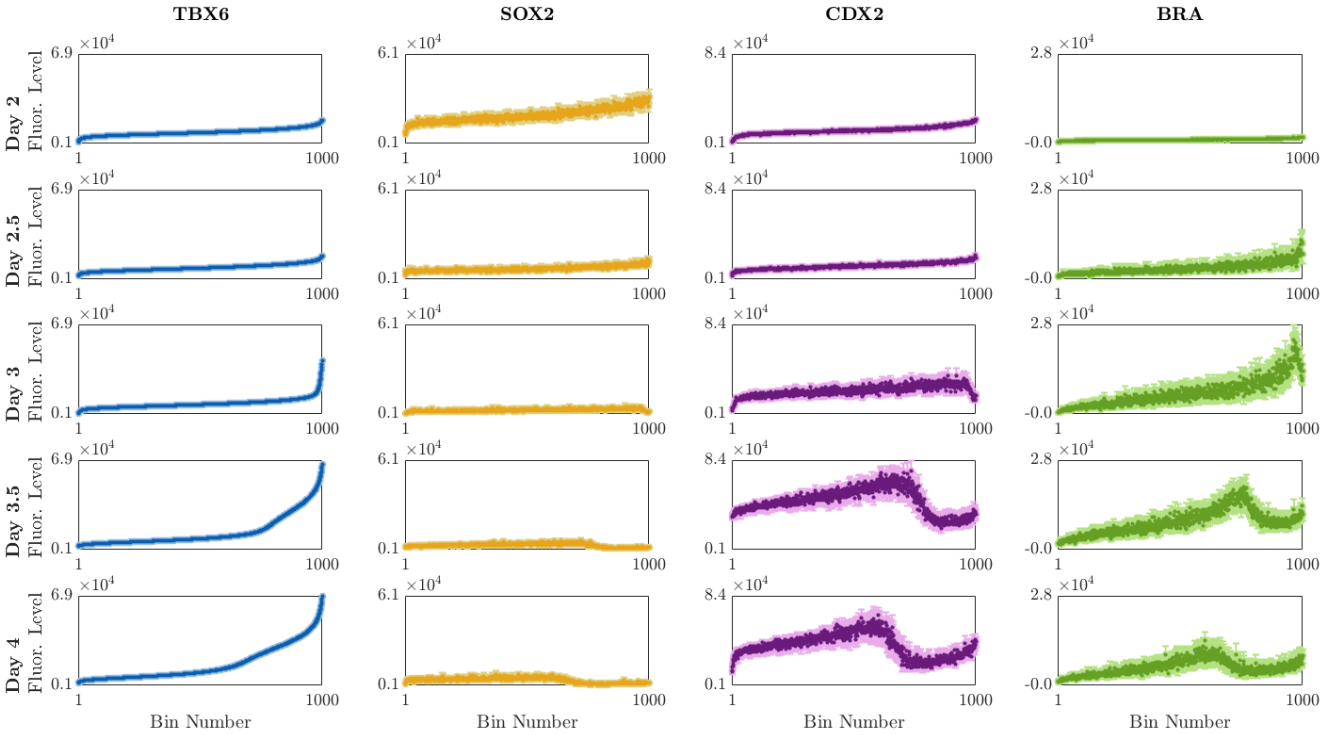


Figure 5.15: *TBX6-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under CHIR condition. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

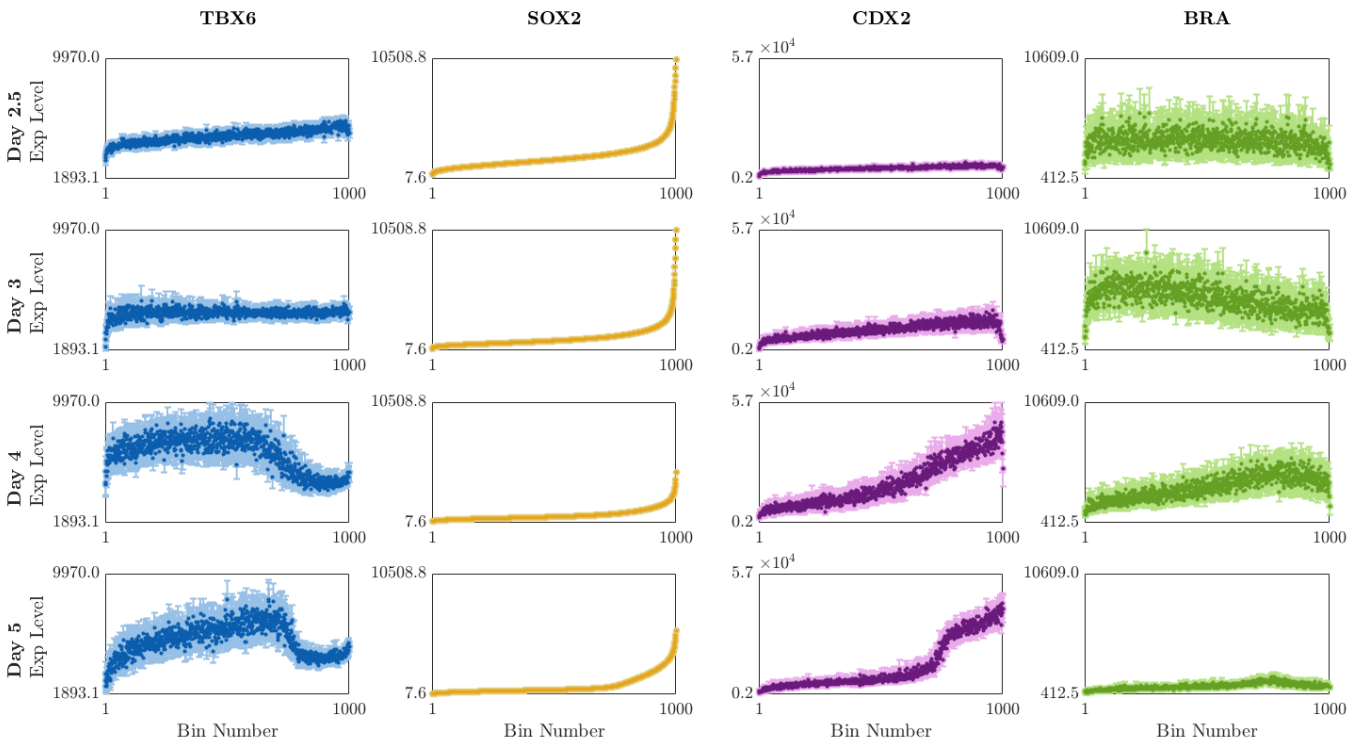


Figure 5.16: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under CHIR condition. The fluorescence levels are taken from Data Set 2. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

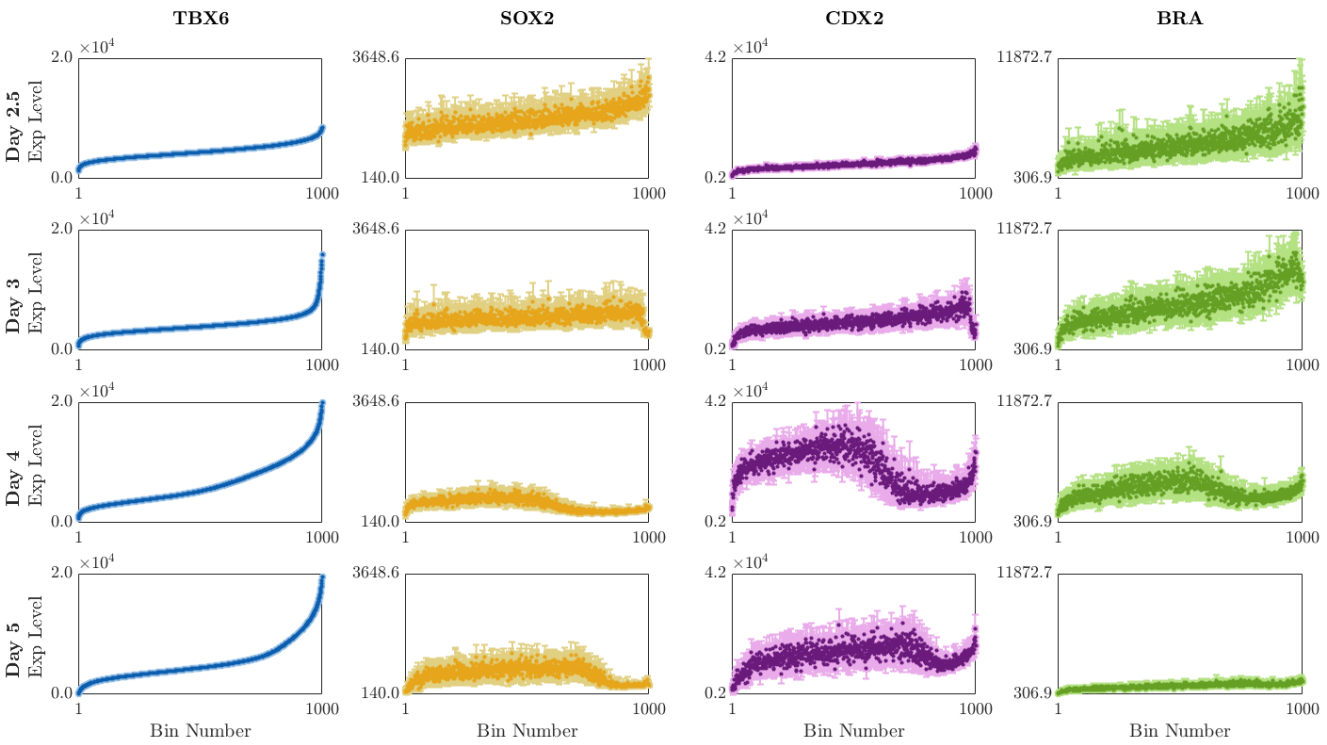


Figure 5.17: *TBX6-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 under CHIR condition. The fluorescence levels are taken from Data Set 2. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

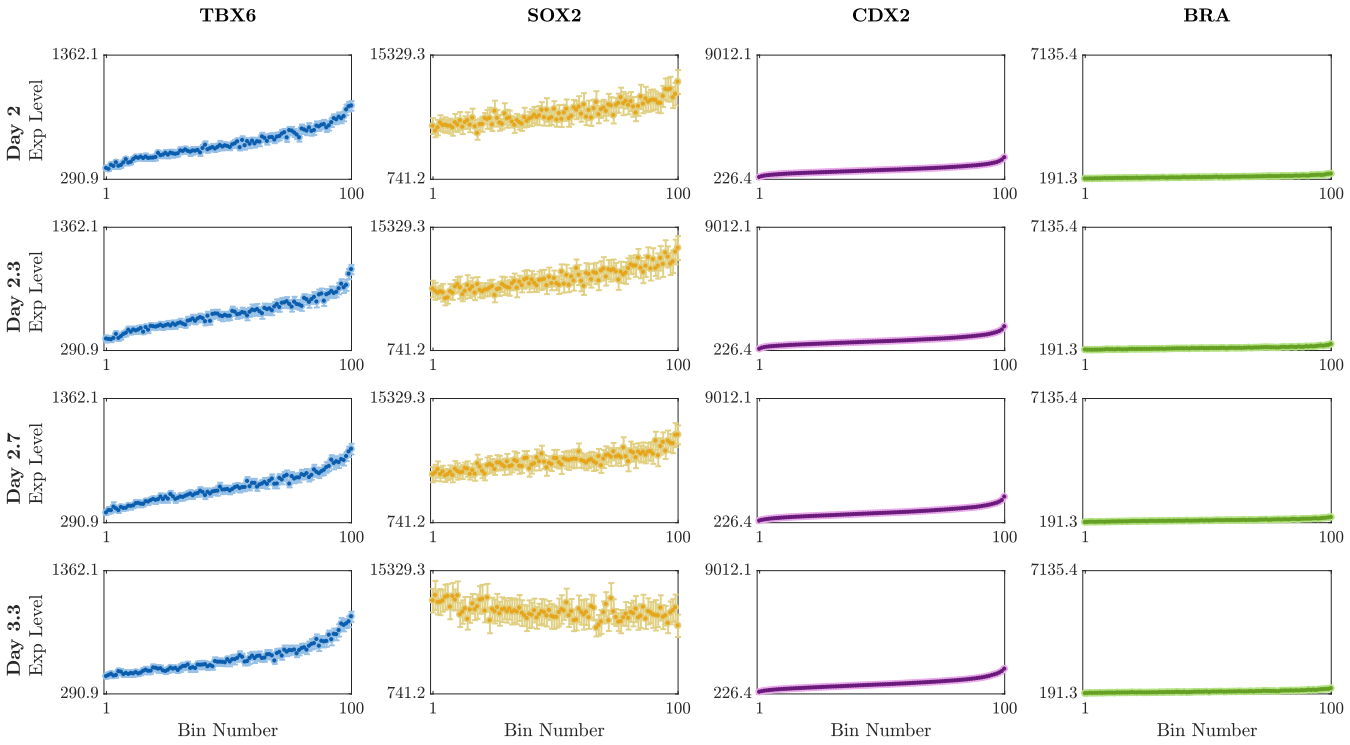


Figure 5.18: *CDX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of WT mESCs under No CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

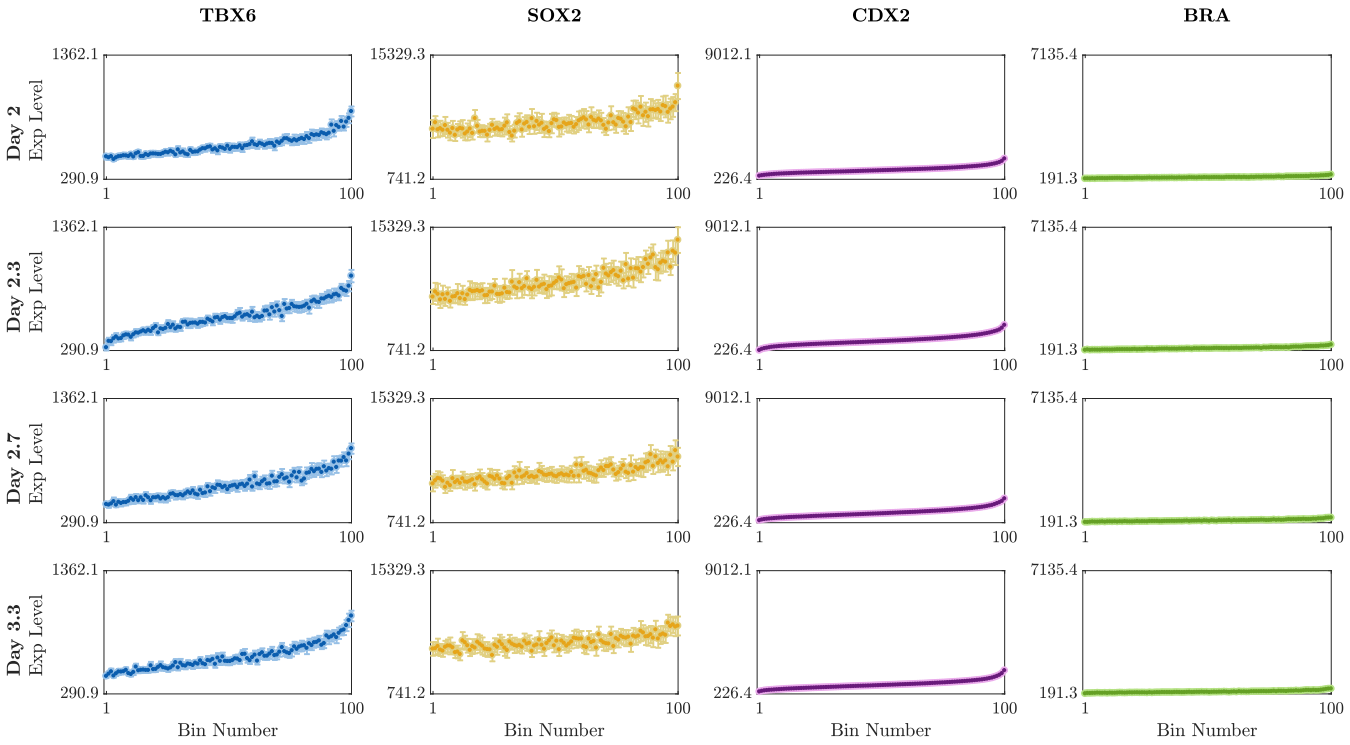


Figure 5.19: *CDX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of Bra$^{-/-}$ mESCs under No CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

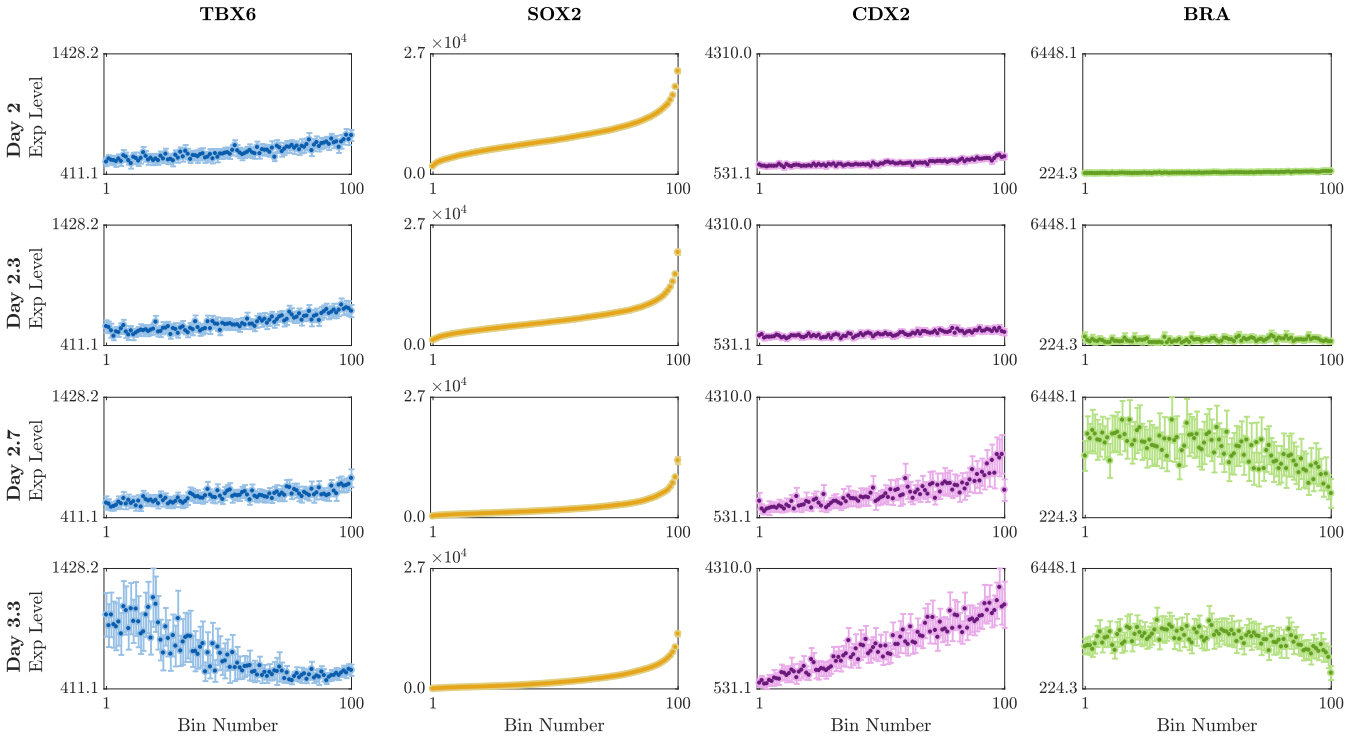


Figure 5.20: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of WT mESCs under CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

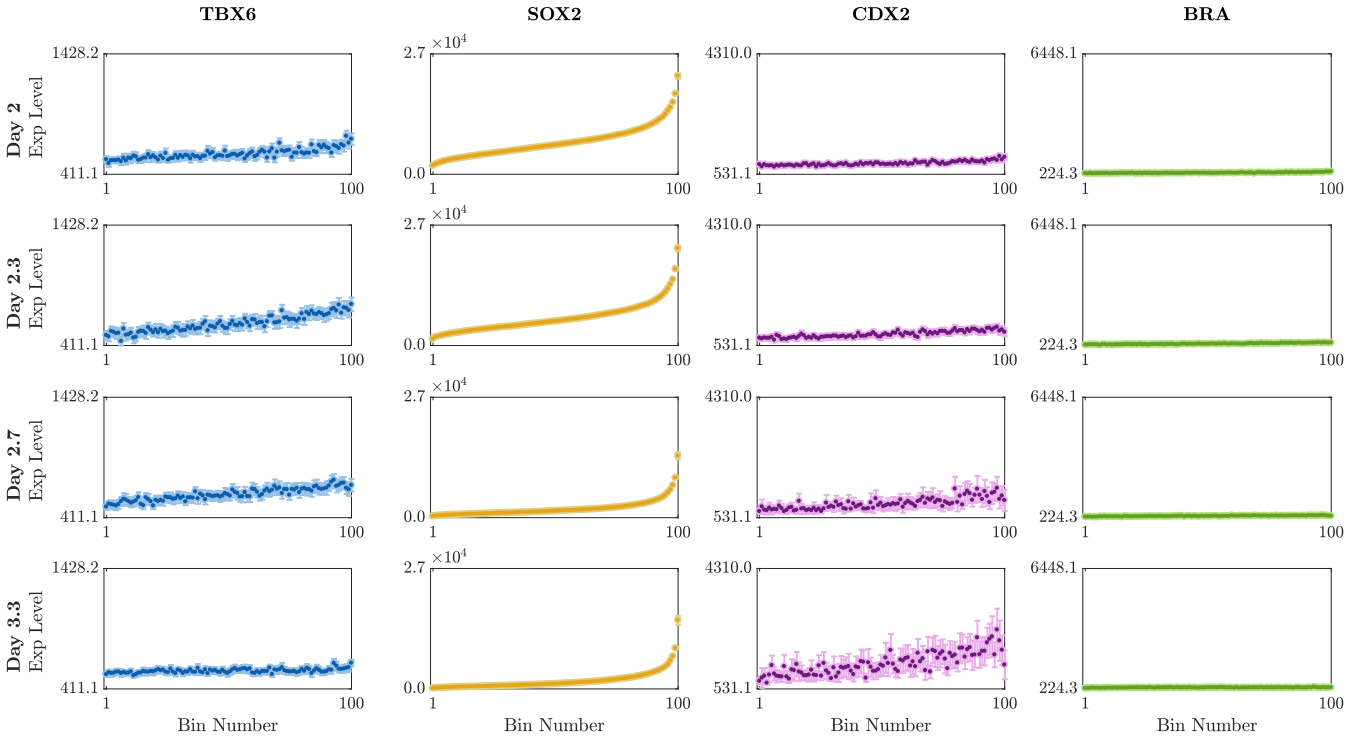


Figure 5.21: *SOX2-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of $Bra^{-/-}$ mESCs under CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

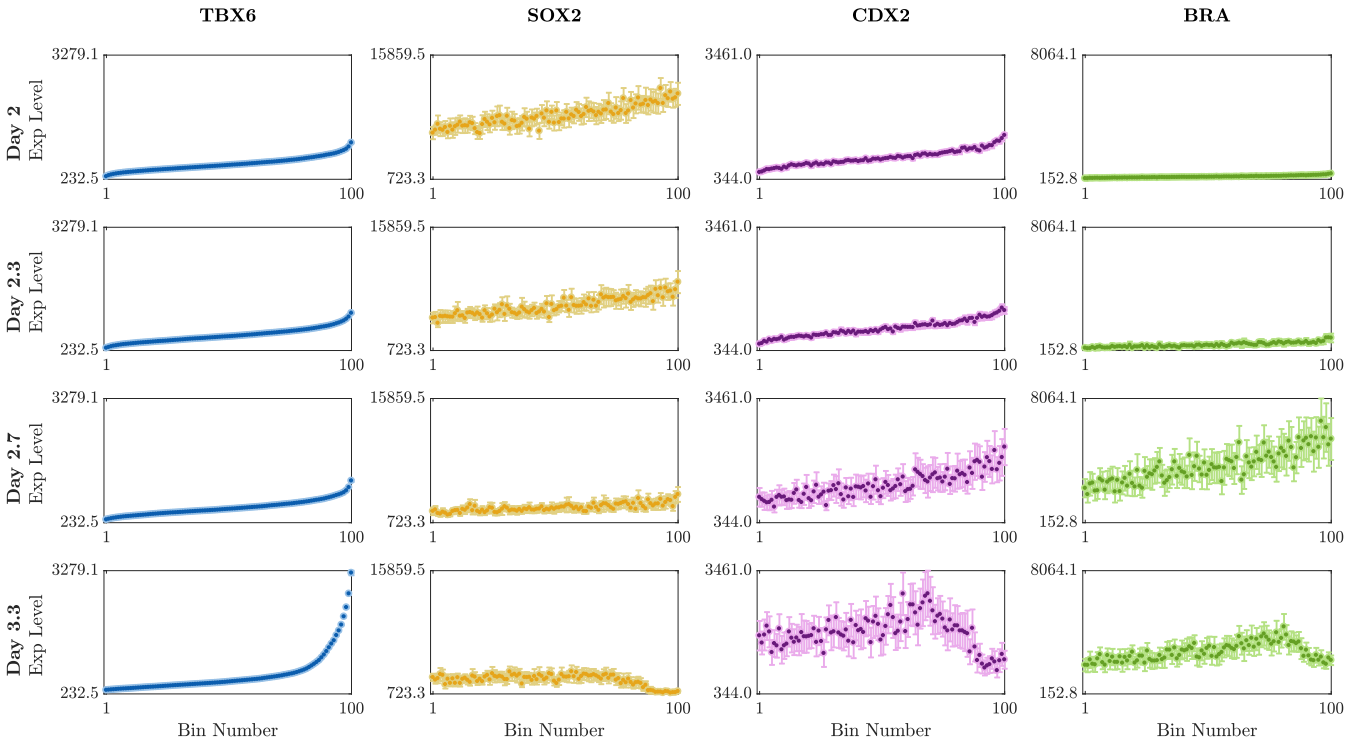


Figure 5.22: *TBX6-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of WT mESCs under CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

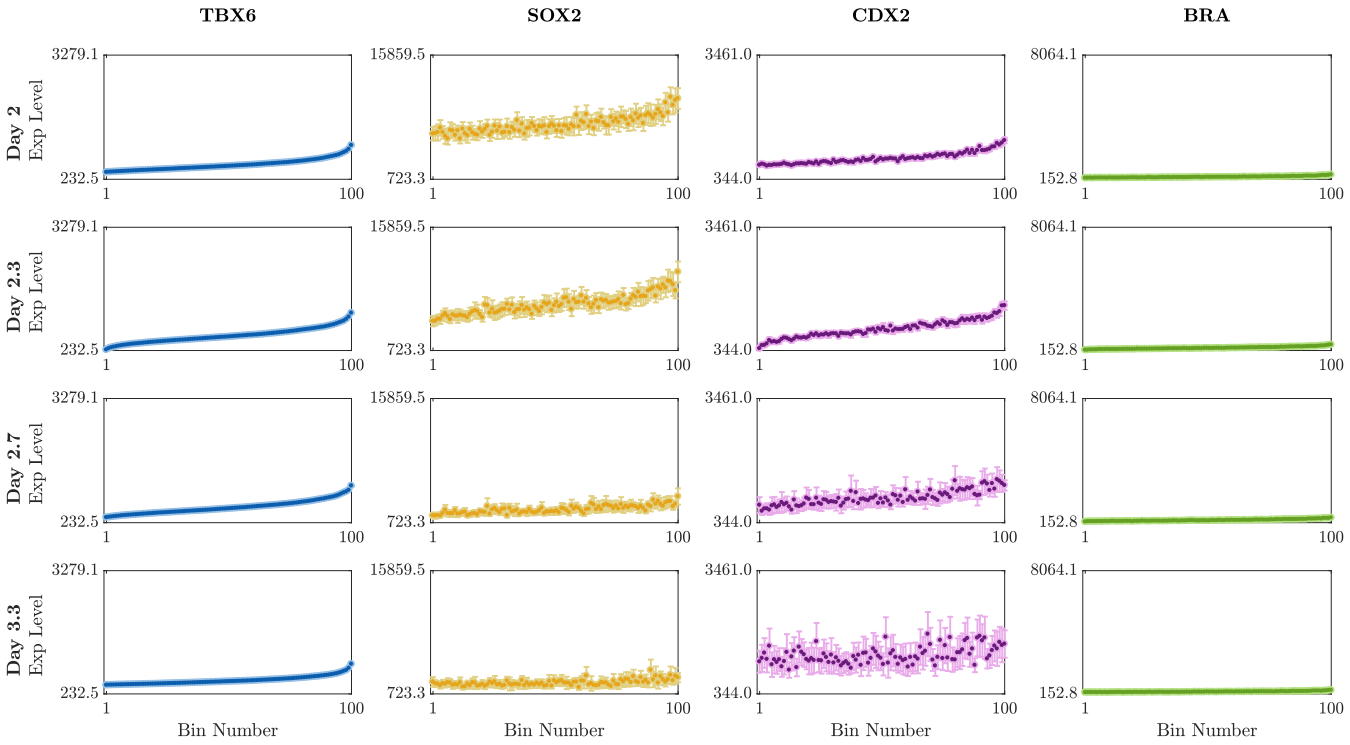


Figure 5.23: *TBX6-ordered 1D GIGLE plots showing the evolution of the protein expression of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) from day 2 to day 4 of Bra$^{-/-}$ mESCs under CHIR condition. The fluorescence levels are taken from Data Set 3. The bin size is of 40 cells per bin. The 95% confidence interval of each bin is represented with error bars in the corresponding light colour.*

In addition to GIGLE plots, violin plots are useful for visualising the dynamics of the genes under different experimental conditions for the population of cells as a whole. Violin plots representing the distributions of the protein levels of each gene in the cells for each day and experimental condition are given in Figure 5.24. The axes have also been fixed for each protein so that the protein levels across conditions are easier to compare.

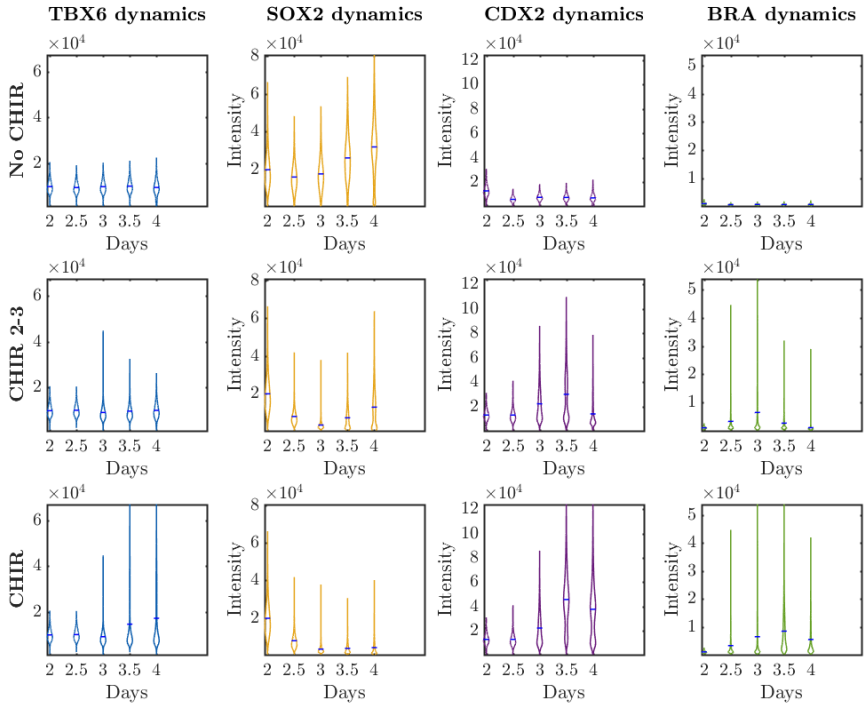


Figure 5.24: *Violin plots of TBX6, SOX2, CDX2 and BRA expression distributions in Data Set 1. First row corresponds to the dynamics of cells that are not exposed to Wnt signalling. Second row of plots corresponds to the dynamics of cells that are exposed to Wnt signalling only from day 2 to day 3. The third row of plots corresponds to the dynamics of cells that are exposed to Wnt signalling from day 2 to day 4. Each column of plots corresponds to the dynamics of TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green). Each violin contains a blue line showing the mean of the protein levels in the sample at the corresponding day for the corresponding gene.*

We can already see in Figure 5.24 that, as expected, the trajectories given by the mean protein expression of the population of cells in time depend on the experimental condition, showing that cells respond differently to different Wnt signalling conditions. Taking advantage of 1D and 4D GIGLE plots we will be able to distinguish different cell types in the samples. We will show that cells developing under the same experimental condition can also follow different developmental pathways,

resulting in two populations of two different cell types at the end of the experiment. Taking advantage of both GIGLE and violin plots, we will be able to understand the cell types that arise under the three experimental conditions studied and also obtain information about the gene expression dynamics of each cell type, which will be very useful for the development of a mathematical model.

As explained in Subsection 5.2.1, mESCs are plated in Fgf and they become **epiblast**-like cells at day two. This epiblast state is characterised by high *Sox2* expression and low expression of the rest of genes (see day 2 in Figure 5.24 or in the 1D GIGLE plots Figures 5.12–5.14). From day 2 onwards, as described in Subsection 5.2.2, cells are exposed to three different experimental conditions for which we will study the differentiation process.

5.3.3.1 Cell differentiation under No CHIR condition

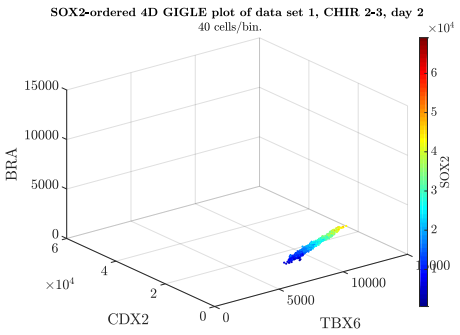
Under this condition, we can see in the 1D GIGLE plots in Figure 5.12 and in the violin plots in Figures 5.24 that the population of cells does not express TBX6, CDX2 and BRA at any of the time points analysed, while SOX2 expression increases in time. In the GIGLE plots we can only distinguish one population of neural ($SOX2^+$) cells. These are, in particular, anterior NPs since they do not express CDX2 in any of the days. This coincides with findings by Gouti et al. (2017) who show that epiblast-like cells that are not exposed to Wnt signalling generate neural precursor cells (SOX2 possitive cells).

5.3.3.2 Cell differentiation under CHIR 2-3 condition

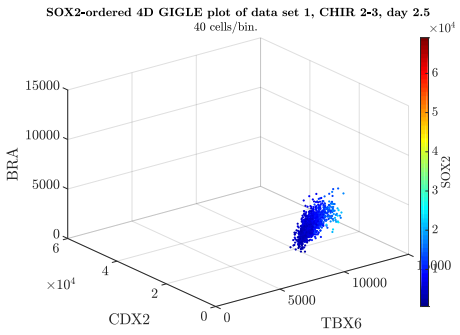
Once epiblast-like cells are exposed to Wnt signalling, we can observe a decrease in SOX2 expression in Figures 5.13 and 5.24. Moreover, a quick increase in BRA expression is also noticeable, with maximal expression at day 3. CDX2 expression also starts to significantly increase at day 3 and reaches its maximum at day 3.5, when its expression starts to decrease. Moreover, TBX6 seems to behave similarly to CDX2, increasing its expression up to day 3 after which it quickly starts to go down.

Figure 5.25 shows GIGLE plots of these samples. At day 3, the population of cells seems to start dividing into two clusters which, by day 4, cluster together again, although a gradient of SOX2 can be seen. It might be noticed in Figure 5.25 that TBX6 and CDX2 are highly correlated at day 2 and 2.5. We suspect that this is a technical artefact coming from the bad efficiency of the antibodies in for both genes, from which the flow cytometer assigns a range of values due to the

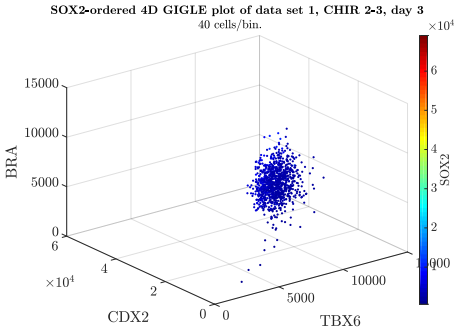
background signal characteristic of the cells, rather than expression of the target protein.



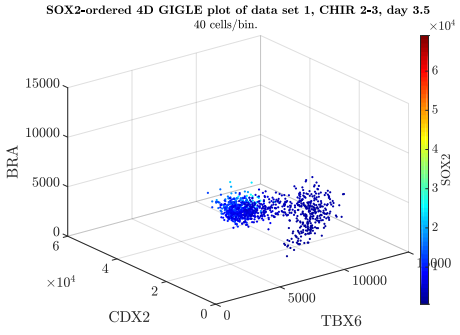
(A) *CHIR 2-3, Day 2*



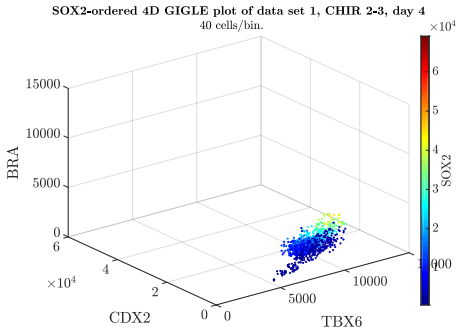
(B) *CHIR 2-3, Day 2.5*



(C) *CHIR 2-3, Day 3*



(D) *CHIR 2-3, Day 3.5*



(E) *CHIR 2-3, Day 4*

Figure 5.25: *SOX2-ordered 4D GIGLE plots showing the evolution of the TBX6, SOX2, CDX2 and BRA expression from day 2 to day 4 under CHIR 2-3 condition. The points are coloured by their relative SOX2 expression and are represented in (TBX6,CDX2,BRA) space. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin, and axes are fixed.*

From these plots we infer that cells reach NMP state between day 2.5 and day 3 if they are exposed to CHIR from day 2, since they co-express BRA and SOX2

in those days (see day 2.5 in Figure 5.13, in agreement with Henrique et al. (2015), Gouti et al. (2014). However, after CHIR is removed at day 3, the expression of all the genes decreases except SOX2 which expression starts to increase. We can see in Figures 5.13, 5.24 and 5.25 that cells end up with a $SOX2^+$ profile by day 4. Therefore, we can say that under CHIR 2-3 condition cells become neural precursor cells, and the dynamics of the gene expression can be observed in the violin plots in Figure 5.24.

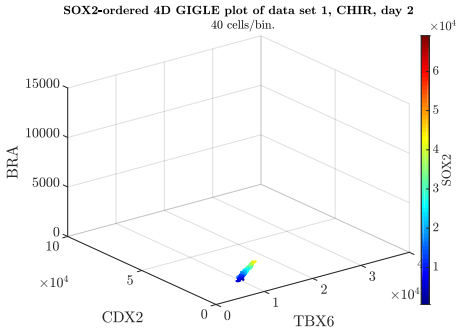
5.3.3.3 Cell differentiation under CHIR condition

As we saw before, Data Set 1 shows that epiblast-like cells become NMP between days 2.5 and 3 if they are exposed to CHIR from day 2. If CHIR is sustained from day 3 onwards, Figures 5.14 and 5.26 clearly show that the population of cells divides into two subpopulations with different gene expression profiles.

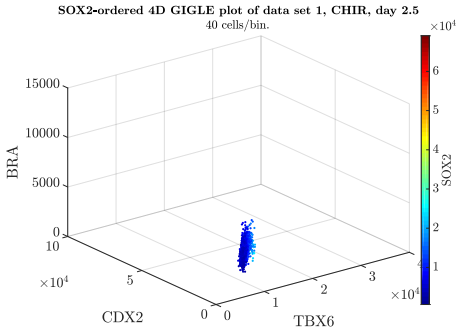
From day 3.5 the bins begin to split into a ($TBX6^+$, BRA^+, $SOX2^-$, $CDX2^-$) population and a ($TBX6^-$, BRA^+, $SOX2^+$, $CDX2^+$) population (see Figures 5.14 and 5.26, day 3.5). The former population downregulates BRA by day 4, becoming TBX6 positive, BRA negative. This is the expression profile corresponding to MPs, as we mentioned in Section 5.1. However, the latter population maintains its BRA expression and upregulates SOX2 and CDX2, so they remain being NMP cells. Hence, under CHIR condition, Data Set 1 shows that the population of epiblast-like cells differentiate into a population of mesodermal progenitors and a population of neuromesodermal progenitors by day 5. We will come back to this NMP population in Section 5.7.

The fact that there are two populations in the CHIR samples means that the violin plots in Figure 5.24 for CHIR condition mix the dynamics of two populations of cells. In order to separate the dynamics of the cells that become MPs from the cells that become NMPs, we decide to take advantage of the clusters obtained from applying the GIGLE method to the data.

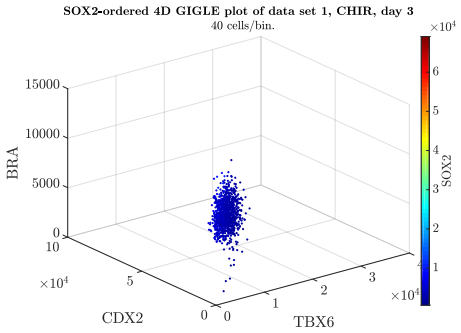
We order the CHIR samples in Data Set 1 with respect to SOX2 expression and bin it into $1,000$ bins. This groups the data into two clear clusters from day 3.5 to day 4, as we can see in Figure 5.26, which allows us to succesfully use k-means clustering in the binned data. We apply **k-means clustering** with $k = 2$ in the binned samples corresponding to days 3.5 and 4, separately, to label the bins as MPs or NMPs. With this clustering technique, we can then produce violin plots for the dynamics of each type of cell separately (see Figure 5.27). We can observe that NMPs express higher SOX2, CDX2 and BRA from day 3.5 than MPs which express higher TBX6 instead, as we saw the 4D GIGLE plots,



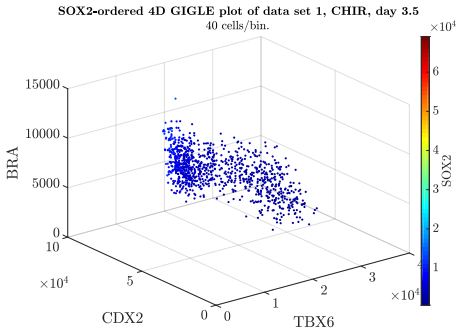
(A) *CHIR, Day 2*



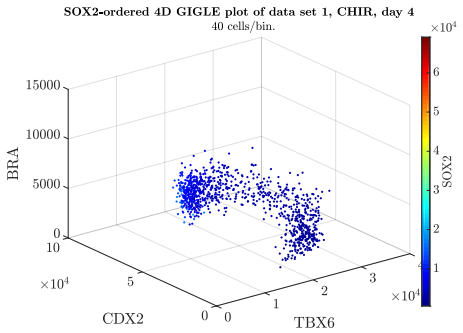
(B) *CHIR, Day 2.5*



(C) *CHIR, Day 3*



(D) *CHIR, Day 3.5*



(E) *CHIR, Day 4*

Figure 5.26: *SOX2-ordered 4D GIGLE plots showing the evolution of the TBX6, SOX2, CDX2 and BRA expression from day 2 to day 4 under CHIR condition. The points are coloured by their relative Sox2 expression and are represented in (TBX6,CDX2,BRA) space. The fluorescence levels are taken from Data Set 1. The bin size is of 40 cells per bin, and axes are fixed.*

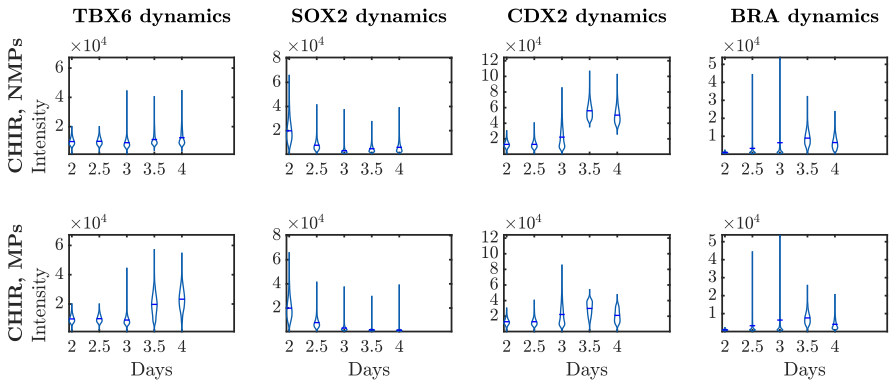


Figure 5.27: *Violin plots of TBX6, SOX2, CDX2 and BRA expression distributions for cells that stay as NMPs (first row) and cells that become MPs (second row) under CHIR condition in Data Set 1. They are obtained by performing k-means clustering after applying GIGLE to samples taken at days 3.5 and 4. Each violin contains a blue line showing the mean of the sample at the corresponding day and for the corresponding gene.*

5.4 Gene Regulatory Network inferred from GIGLE

We now study the interactions between the genes *Tbx6*, *Sox2*, *Cdx2*, *Bra* and the Wnt and RA signals that control the differentiation of mESCs into NPs, NMPs or MPs by taking advantage of the GIGLE plots. We propose that the gene regulatory network is the one given in Figure 5.28. In this section we will provide evidence of each interaction.

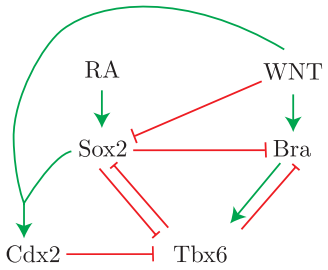


Figure 5.28: *Proposed gene regulatory network that controls the specification of mouse epiblast progenitors into neuromesodermal progenitors, neural and mesodermal progenitors. Green arrows represent activation of gene expression of the downstream node, red arrows represent inhibition of gene expression of the downstream node.*

Wnt upregulates *Bra* Yamaguchi et al. (1999) show that *Bra* is a direct target of Wnt, being downregulated in Wnt mutants. Moreover, we can observe this induction in the GIGLE plots. Comparing Figure 5.12 where WT epiblast mESCs have not been exposed to CHIR and Figure 5.14 where they have been exposed to CHIR every day, we can see that the BRA fluorescence levels

are very low if CHIR is not present. Also, comparing the 1D GIGLE plots in Figure 5.14 where cells have been exposed to CHIR every day and Figure 5.13 where cells have been exposed to CHIR only from day 2 to day 3, one can notice a stronger downregulation of *Bra* on days 3.5 and 4 in the latter conditions. We therefore conclude that there is an induction of *Bra* due to Wnt signalling.

Wnt downregulates *Sox2* Comparing Figure 5.22 where WT mESCs have been exposed to CHIR every day and Figure 5.23 where $Bra^{-/-}$ mESCs have been exposed to CHIR every day too, we can see that the SOX2 fluorescence levels decrease each day, even if the rest of the proteins levels remain the same (except BRA and TBX6 which are not expressed in *Bra* mutant). However, if CHIR is not added to the media, as it is the case in Figures 5.18 and Figures 5.19 where WT and $Bra^{-/-}$ mESCs have not been exposed to CHIR, we can observe that the fluorescence levels of *Sox2* remain high or even increase in time. We believe that these are evidences for Wnt downregulating *Sox2* independently of *Bra* and *Tbx6*.

Both Wnt and *Sox2* are necessary for *Cdx2* expression Nordström et al. (2006) and van de Ven et al. (2011) show a regulation of *Cdx2* by Wnt in chick and mouse embryos respectively. We can observe in the Figures where CHIR is not added to the culture that *Cdx2* is not expressed (see Figure 5.12 as an example). We can also observe that, when CHIR is added to the culture (see Figure 5.17 for example) the expression of *Cdx2* increases in time and by day 4 or 5 it is highly expressed but only in cells that express *Sox2*. Furthermore, CDX2 induction by CHIR is compromised in $Sox2^{-/-}$ progenitors (personal communication from Dr. Blassberg). We infer from this that both Wnt and *Sox2* expression are necessary for *Cdx2* expression.

Cdx2* downregulates *Tbx6 Our collaborators have some evidence that *Cdx2* can repress *Tbx6*. We also see strong anticorrelations between CDX2 and TBX6 levels in Figures 5.14, 5.16 and 5.20 from day 3.5. Therefore we decide to add this link to the network.

***Bra* is necessary for *Tbx6* expression** Comparing Figures 5.20 and 5.21 we can notice that *Tbx6* is not expressed in $Bra^{-/-}$ cells and it is expressed in WT cells at day 3.3 after an upregulation of *Bra* at day 2.7.

Tbx6* downregulates *Bra We can see in Figures 5.15 and 5.17 that *Tbx6* expression is highly anticorrelated with *Bra* expression from day 3, when *Tbx6*

expression starts. We see that in bins where *Tbx6* expression is high, *Bra* expression is very low.

***Tbx6* upregulates itself** We do not have evidence for this, actually *Tbx6* positive cells that become MPs downregulate it after day 5. However, we need to add this link to the network to make the MP state stable, otherwise we will obtain oscillations in *Tbx6* expression as it downregulates *Bra*.

Sox2* inhibits *Bra Thomson et al. (2011) show that cells with high constitutive *Sox2* expression[1] did not express *Bra*.

RA activates *Sox2* We obtain this link from Gouti et al. (2017). They show that SOX2 expression is abolished in Aldh1a2$^{-/-}$ mESCs (mutants of the RA-producing enzyme Aldh1a2).

***Sox2* autoregulates itself** There is evidence of this link in human ESCs (Boyer et al. 2005) and our collaborators have (unpublished) data that show that this happens in mouse ESCs too.

Tbx6* downregulates *Sox2 Takemoto et al. (2011) show that, in mouse embryos, *Tbx6* represses *Sox2* by inactivating a *Sox2* enhancer. Moreover, Gouti et al. (2017) show that this is also true in mESCs by comparing *Sox2* expression in WT mESCs and in a *Tbx6*$^{-/-}$ cell line.

Sox2* downregulates *Tbx6 We can see also in Figures 5.14, 5.16 and 5.20 that, from day 3.5, there is a strong anticorrelation between *Sox2* and *Tbx6*. We can observe that *Tbx6* is expressed only in bins where *Sox2* and *Cdx2* are not expressed. Figure 5.13 shows that, *Tbx6* is anticorrelated with *Sox2* at day 4 even if *Cdx2* is not expressed, so *Cdx2* is not necessary for the downregulation of *Tbx6*. This effect could also be due to *Sox2* repression by *Tbx6*, however we decide to add this node to the network as it is in the network proposed by Gouti et al. (2017) and we cannot provide data against it.

Unlike the GRN proposed in Gouti et al. (2017), we do not think that *Bra* downregulates *Sox2*. Comparing the *Sox2* levels under CHIR condition in Figure 5.22, where *Bra* is present, and the ones in Figure 5.23, where cells are *Bra*$^{-/-}$, are practically the same (with a downregulation of *Sox2* on day 3.3 which can be explained by the upregulation of *Tbx6*).

[1] A constitutive gene is a gene that is transcribed continually in contrast to a facultative gene which is only transcribed as needed

We will show that with this gene regulatory network we can reproduce most of the features in Data Set 1 from day 2 to day 4 that we studied in Subsection 5.3.3.

5.5 Deterministic mathematical model

We develop a deterministic mathematical model for the gene regulatory network proposed in 5.28. The variables represent the mRNA concentration of *Tbx6* (x_T), *Sox2* (x_S), *Cdx2* (x_C) and *Bra* (x_B). We model the corresponding protein levels y_T, y_S, y_C and y_B as a delayed version of the mRNA using the linear chain trick (we refer the reader to Fall (2002) for a derivation of the linear chain trick). The system has two main input signals which are Wnt (W) and RA (RA). We also model the response of a cell to the input signals Wnt (W_R) and RA (RA_R) with a delay, obtaining the following equations:

$$
\begin{aligned}
\frac{dW_R}{dt} &= \alpha_W(W - W_R)\\
\frac{dRA_R}{dt} &= \alpha_{RA}(RA - RA_R)\\
\frac{dx_T}{dt} &= v_T\left[\frac{(\theta_{BT}y_B)^{n_{BT}} + (\theta_{TT}y_T)^{n_{TT}}}{1 + (\theta_{BT}y_B)^{n_{BT}} + (\theta_{TT}y_T)^{n_{TT}}}\right]\left[\frac{1}{1 + (\theta_{ST}y_S)^{n_{ST}}}\right]\left[\frac{1}{1 + (\theta_{CT}y_C)^{n_{CT}}}\right] - k_T x_T\\
\frac{dy_T}{dt} &= \alpha_T(x_T - y_T)\\
\frac{dx_S}{dt} &= v_S\left[\frac{(\theta_{RS}RA_R)^{n_{RS}} + (\theta_{SS}y_S)^{n_{SS}}}{1 + (\theta_{RS}RA_R)^{n_{RS}} + (\theta_{SS}y_S)^{n_{SS}}}\right]\left[\frac{1}{1 + (\theta_{WS}W_R)^{n_{WS}}}\right]\left[\frac{1}{1 + (\theta_{TS}y_T)^{n_{TS}}}\right] - k_S x_S\\
\frac{dy_S}{dt} &= \alpha_S(x_S - y_S)\\
\frac{dx_C}{dt} &= v_C\left[\frac{(\theta_{WC}W_R)^{n_{WC}}}{1 + (\theta_{WC}W_R)^{n_{WC}}}\right]\left[\frac{(\theta_{SC}y_S)^{n_{SC}}}{1 + (\theta_{SC}y_S)^{n_{SC}}}\right] - k_C x_C\\
\frac{dy_C}{dt} &= \alpha_C(x_C - y_C)\\
\frac{dx_B}{dt} &= v_B\left[\frac{(\theta_{WB}W_R)^{n_{WB}}}{1 + (\theta_{WB}W_R)^{n_{WB}}}\right]\left[\frac{1}{1 + (\theta_{TB}y_T)^{n_{TB}}}\right]\left[\frac{1}{1 + (\theta_{SB}y_S)^{n_{SB}}}\right] - k_B x_B\\
\frac{dy_B}{dt} &= \alpha_B(x_B - y_B).
\end{aligned}
\tag{5.1}
$$

The regulations affecting the synthesis of the genes are described by Hill functions. For example, in the equation describing the evolution of *Tbx6* mRNA, x_T, the first term correspond to the synthesis of *Tbx6* activated by BRA and TBX6 itself. Since at least one of the two activations is necessary, they appear as addends in the same Hill function. The second and third term correspond to the inhibitory influence of SOX2 and CDX2 respectively, assumed to affect the activation and thus they appear as multiplicative terms. The Hill coeficients are chosen to take values 1, 2 or 3. The degradations of the genes and proteins are described by the negative term of the equations modelling the mRNA and protein evolution. For simplicity, we assume that these reactions follow first order kinetics. The delays in translation are controlled by the parameters α. Tables 5.1 and 5.2 contain the list of parameters with the corresponding descriptions and values.

5.5.1 Stability of the model

Taking into account the analysis developed in Subsection 5.3.3, the system needs to be bistable under CHIR conditions. Under this condition, Wnt signalling is high and RA signalling is low (since this signal is not added to the media but it is produced by the cells themselves). The two attractors must also correspond to the expression profiles of NMP cells ($SOX2^+$, $CDX2^+$ and BRA^+) and mesodermal progenitor cells ($TBX6^+$).

Indeed, the system of equations in 5.1 is bistable for a range of values of the parameters W and RA, when the rest of parameters take the values given in Tables 5.1 and 5.2. The stable critical points have coordinates corresponding to the desired gene expression profiles and W and RA act as bifurcation parameters (see Figure 5.29).

The system is bistable for a range of values of W and RA, with the critical points representing NMP and mesodermal progenitor profiles. By increasing W or decreasing RA, the system becomes monostable with the critical point representing mesodermal progenitor profile. On the other hand, by decreasing W or increasing RA, the critical point representing mesodermal progenitor cells bifurcates and the system becomes monostable, the only attractor corresponding to NMP gene expression profile.

Table 5.1: *Table of parameters of the deterministic model in Equation 5.1 for the proposed gene regulatory network controlling neural and mesoderm specification during vertebrate trunk development.*

Parameter name	Parameter meaning	Value
WNT	WNT (CHIR) signal	
RA	RA signal	
v_T	Transcription rate of *Tbx6*	0.8
v_S	Transcription rate of *Sox2*	0.5
v_C	Transcription rate of *Cdx2*	1
v_B	Transcription rate of *Bra*	1
θ_{BT}	Efficiency of *Tbx6* activation by BRA	2
n_{BT}	Hill coefficient of *Tbx6* activation by BRA	1
θ_{TT}	Efficiency of *Tbx6* autoactivation	5
n_{TT}	Hill coefficient of *Tbx6* autoactivation	2
θ_{ST}	Efficiency of *Tbx6* repression by SOX2	5
n_{ST}	Hill coefficient of *Tbx6* repression by SOX2	2
θ_{CT}	Efficiency of *Tbx6* repression by CDX2	2
n_{CT}	Hill coefficient of *Tbx6* repression by CDX2	2
θ_{RS}	Efficiency of *Sox2* activation by RA	5
n_{RS}	Hill coefficient of *Sox2* activation by RA	3
c_{SS}	Efficiency of *Sox2* autoactivation	3
n_{SS}	Hill coefficient of *Sox2* autoactivation	2
θ_{WS}	Efficiency of *Sox2* repression by WNT	3
n_{WS}	Hill coefficient of *Sox2* repression by WNT	2
θ_{TS}	Efficiency of *Sox2* repression by TBX6	1

Table 5.2: *Table of parameters of the deterministic model in Equation 5.1 for the proposed gene regulatory network controlling neural and mesoderm specification during vertebrate trunk development (Continuation).*

Parameter name	Parameter meaning	Value
n_{TS}	Hill coefficient of *Sox2* repression by TBX6	2
θ_{WC}	Efficiency of *Cdx2* activation by WNT	1
n_{WC}	Hill coefficient of *Cdx2* activation by WNT	1
θ_{SC}	Efficiency of *Cdx2* repression by SOX2	8
n_{SC}	Hill coefficient of *Cdx2* repression by SOX2	1
θ_{WB}	Efficiency of *Bra* activation by WNT	1
n_{WB}	Hill coefficient of *Bra* activation by WNT	1
θ_{TB}	Efficiency of *Bra* repression by TBX6	3
n_{TB}	Hill coefficient of *Bra* repression by TBX6	2
θ_{SB}	Efficiency of *Bra* repression by SOX2	1
n_{SB}	Hill coefficient of *Bra* repression by SOX2	2
k_T	Degradation rate of TBX6	1
k_S	Degradation rate of SOX2	0.5
k_C	Degradation rate of CDX2	0.8
k_B	Degradation rate of BRA	1
α_W	Imput rate of Wnt	0.5
α_{RA}	Imput rate of RA	1
α_T	Translation rate of *Tbx6* and degradation rate of TBX6	0.8
α_S	Translation rate of *Sox2* and degradation rate of SOX2	0.8
α_C	Translation rate of *Cdx2* and degradation rate of CDX2	0.5
α_B	Translation rate of *Bra* and degradation rate of BRA	0.5

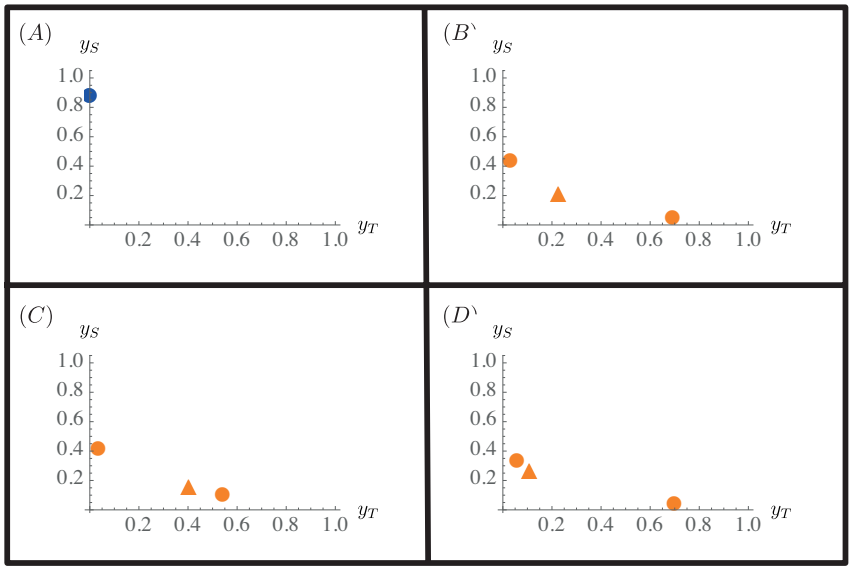


Figure 5.29: *Example of configurations of critical points that can be obtained from model in Equation 5.1 by changing the values of the parameters W and RA. Stable critical points are represented as circles, saddle points are represented as triangles. They are coloured in blue if all the eigenvalues of the jacobian at the point are real, orange otherwise. (A) This configuration is obtained by setting $W = 0$, $RA = 0.09$ (which we regard as null WNT and low RA). There is only one critical point for these parameter values, with coordinates $(x_T, x_S, x_C, x_B) = (0, 0.87, 0, 0)$. Therefore, under the conditions of null Wnt signal and low RA signal, the model is monostable and the critical point has coordinates corresponding to the anterior neural progenitor profile.(B) This configuration is obtained by setting $W = 0.23$, $RA = 0.09$ (regarded as high WNT and low RA). There are three critical points: a saddle point with coordinates $(0.23, 0.19, 0.14, 0.12)$, a stable critical point with coordinates $(0.03, 0.43, 0.18, 0.16)$ (which corresponds to NMP profile), and another stable critical point with coordinates $(0.69, 0.05, 0.06, 0.04)$ (which corresponds to mesodermal progenitor profile). (C) This configuration is obtained with $W = 0.16$ and $RA = 0.09$. We can see that by decreasing W we can make the attractor representing mesodermal progenitor and the saddle bifurcate. The same happens if $W = 0.24$ and RA is increased to 0.106 (D) This configuration is obtained by setting $W = 0.25$ $RA = 0.09$. By increasing W (or decreasing RA) the attractor representing the NMP state can also bifurcate.*

5.5.2 Deterministic solutions under different Wnt and RA conditions

We numerically solved the system in Equation 5.1 from $t_0 = 0$ to $t_1 = 20$ with the Matlab **ode15s** solver, with parameter values given in Tables 5.1 and 5.2 under different Wnt and RA conditions to obtain the dynamics of the protein expression. In particular, we would like to study the solutions under No CHIR (null Wnt, low RA), CHIR 2-3 (high Wnt from day 2 to day 3 and null from day 3, low RA) and CHIR (high Wnt, low RA) conditions.

In order to translate days in the experiment into time t in our model, we di-

vide the interval $[t_0, t_1]$ into four subintervals, the extremes of which will be mapped to the days. This way $t = 0$ corresponds to day 2, $t = 5$ corresponds to day 2.5, $t = 10$ corresponds to day 3, $t = 15$ corresponds to day 3.5 and $t = 20$ corresponds to day 4.

In the following subsections we will provide the deterministic solutions obtained from our model under each experimental condition.

5.5.2.1 No CHIR condition

This experimental condition can be translated as $W = 0$ and $RA = 0.09$ in our model. As we see in Figure 5.29, under this parameter values only an attractor corresponding to anterior neural progenitor is present. Regarding the initial conditions for the system, we choose 0 as initial mRNA levels as well as the initial values of W_R and RA_R. As we saw in Section 5.3.3, epiblast cells are characterised by being just SOX2 positive, therefore we choose low values for the initial TBX6, CDX2 and BRA levels, and a high value for the initial SOX2 level. In particular, we set $y_T(0) = y_B(0) = y_C(0) = 0.001$, and $y_S(0) = 0.3$. Under these conditions, we obtain the solution given in Figure 5.30.

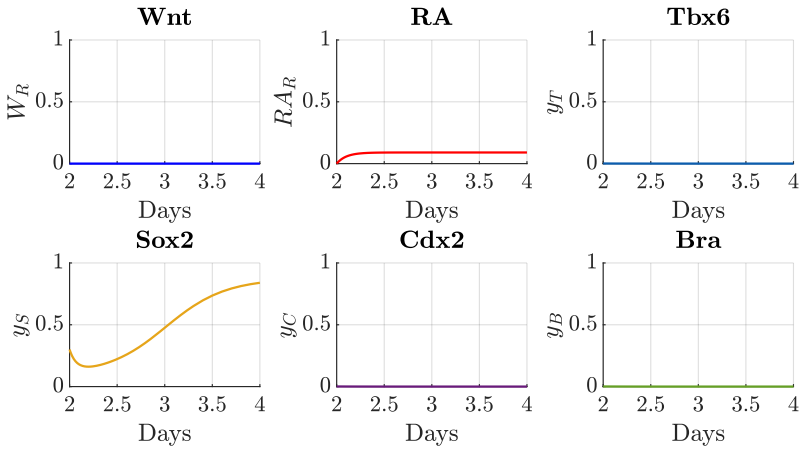


Figure 5.30: *Deterministic solution obtained for W_R and RA_R as well as the TBX6, SOX2, CDX2 and BRA protein levels under No CHIR condition ($W = 0$, $RA = 0.09$). The x-axis corresponding to t has been changed so that it represents days, as explained in the main text.*

The trajectories of the protein levels in Figure 5.30 converge to the point $(y_T, y_S, y_C, y_B) = (0, 0.87, 0, 0)$, that represents the $SOX2^+$ neural progenitor profile, as in the experiments. They resemble the trajectories of the means in the No CHIR violin plots in Figure 5.24. As in the experiments, only SOX2 is expressed

under this contition and cells become NPs. It might be noticeable that SOX2 levels increase slightly faster than in the experimental data, we will discuss this in Section 5.7.

5.5.2.2 CHIR 2-3 condition

This experimental condition is translated as $W = 0.24, RA = 0.09$ from $t = 0$ (day 2) to $t = 10$ (day 3) and $W = 0, RA = 0.09$ from $t = 10$ (day 3) to $t = 20$ (day 4) in our model.

Under this regime, the system will have two attractors from $t = 0$ to $t = 10$ when $W = 0.24$ while it will contain just one attractor corresponding to NP from $t = 15$ to $t = 30$ when $W = 0$ (see Figure 5.29). We choose the same initial conditions as in the No CHIR condition, i.e. $y_T(0) = y_B(0) = y_C(0) = 0.001$, and $y_S(0) = 0.3$. Under these conditions, we obtain the solution given in Figure 5.31.

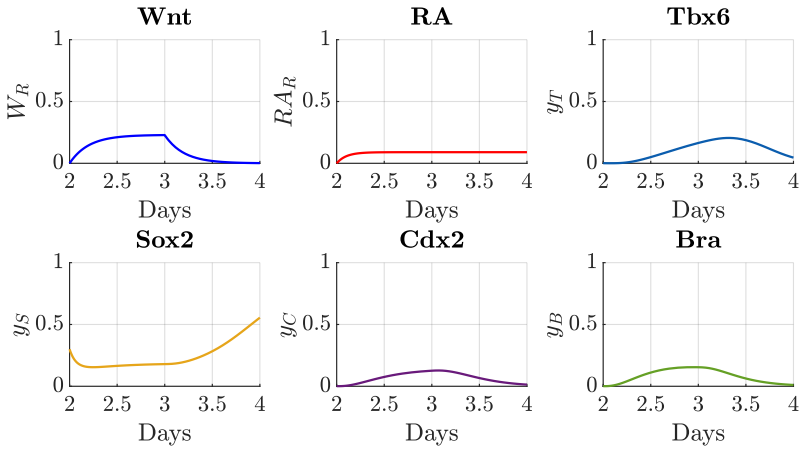


Figure 5.31: *Deterministic solution obtained for W_R and RA_R as well as the TBX6, SOX2, CDX2 and BRA protein levels under CHIR 2-3 condition ($W = 0.24, RA = 0.09$ from $t = 0$ to $t = 10$ and $W = 0, RA = 0.09$ from $t = 10$ to $t = 20$). The x-axis corresponding to t has been changed so that it represents days, as explained in the main text.*

The trajectories of the protein levels in Figure 5.31 look qualitatively similar to the trajectories of the means in the CHIR 2-3 violin plots in Figure 5.24. The trajectory converge to the point $(y_T, y_S, y_C, y_B) = (0, 0.87, 0, 0)$, that represents the SOX2$^+$ neural progenitor profile, but this time it goes through an NMP state first with SOX2, CDX2 and BRA co-expression between days 2.5 and 3, as in the experiments. On the other hand, we see that SOX2 expression does not decrease as much as in the experiments from day 2 to day 3. Again, we will discuss this in Section 5.7.

5.5.2.3 CHIR condition

This experimental condition is translated as $W = 0.24, RA = 0.09$ at all times in our model.

Under this condition, the system will have two attractors corresponding to NMP and mesoderm progenitors (see Figure 5.29). We believe that the fact that two populations of cells are obtained under the same experimental condition is mainly due to the variablility of SOX2 initial levels in the epiblast cells. In order to find the two trajectories corresponding to differentiation into NMP and differentiation into mesodermal progenitor, we choose two initial conditions, which have the same initial TBX6, CDX2 and BRA levels as before ($y_T(0) = y_B(0) = y_C(0) = 0.001$) but differ in the initial SOX2 level. For the trajectory corresponding to NMP differentiation we choose a high SOX2 level $y_S(0) = 0.5$. On the other hand, for the trajectory corresponding to differentiation into mesodermal progenitor we choose a low initial SOX2 level $y_S(0) = 0.1$. Under these conditions, we obtain the solutions given in Figures 5.32 and 5.33.

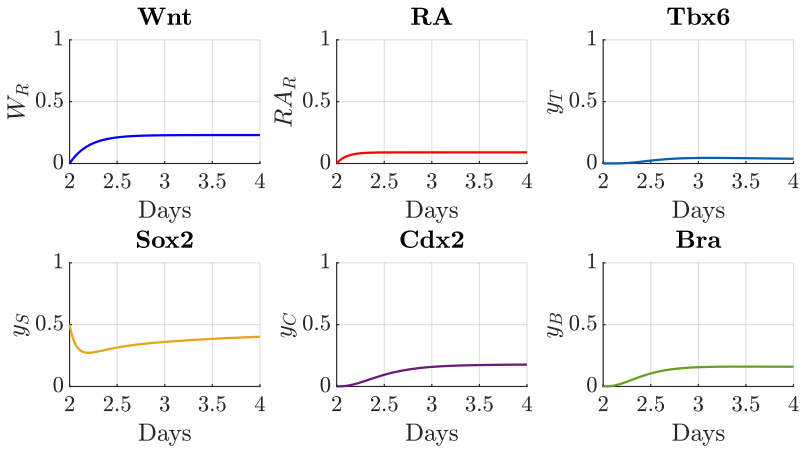


Figure 5.32: *Deterministic solution obtained for W_R and RA_R as well as the TBX6, SOX2, CDX2 and BRA protein levels under CHIR condition for an NMP cell ($W = 0.24, RA = 0.09$, $y_S(0) = 0.5$). The x-axis corresponding to t has been changed so that it represents days, as explained in the main text.*

The trajectories of the protein levels for NMP and MP differentiation in Figures 5.32 and 5.33 resemble the trajectories of the means in the CHIR violin plots in Figure 5.27. In the NMP case, the trajectories converge to a CDX2 and BRA positive profile with a lower but still positive SOX2 level, as in the experiments. The solution converges to the point $(y_T, y_S, y_C, y_B) = (0.03, 0.43, 0.18, 0.16)$, that represents the SOX2$^+$, CDX2$^+$, BRA$^+$ NMP profile. In the MP case, the trajectory

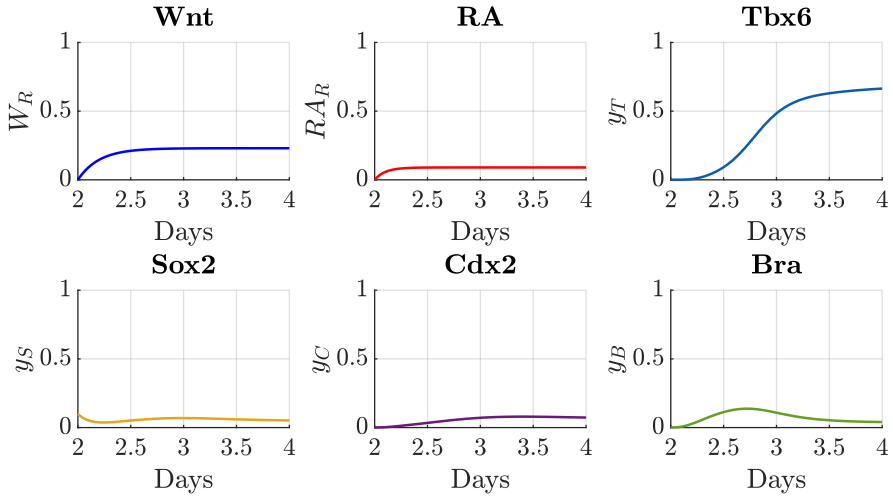


Figure 5.33: *Deterministic solution obtained for W_R and RA_R as well as the TBX6, SOX2, CDX2 and BRA protein levels under CHIR condition for an MP cell ($W = 0.24, RA = 0.09$, $x_S(0) = 0.1$). The x-axis corresponding to t has been changed so that it represents days, as explained in the main text.*

goes through a point of high BRA expression between days 2.5 and 3, which level is then decreased. The TBX6 level increases significantly between days 2.5 and 3 as in the experimental data. The solution converges to the point $(y_T, y_S, y_C, y_B) = (0.69, 0.05, 0.06, 0.04)$, that represents the TBX6$^+$ mesodermal progenitor profile, as we aimed for.

5.6 Stochastic mathematical model

We also develop a stochastic mathematical model to simulate the differentiation of a population of epiblast cells into neural progenitors, NMP or mesodermal progenitors under No CHIR, CHIR 2-3 or CHIR conditions. This way we can more easily test the results of the model by comparing them to the noisy data.

Following the approach in Gonze et al. (2003) we write a stochastic version of the model in Equation 5.1. As we explained in in Section 2.2, in order to obtain the stochastic model we need to define the set of reactions that define the evolution of the system, i.e. the reaction propensities and the stoichiometry matrix. Table 5.3 shows the series of reactions and their corresponding propensities. In this case, the system comprises $n = 10$ species and $m = 20$ reactions.

To derive the propensities $\omega_j(\boldsymbol{Y})$ we use the relation $\omega_j(\boldsymbol{Y}) = \Omega u_j(\boldsymbol{Y}/\Omega) = \Omega u_j(\boldsymbol{X})$ as explained in Section 2.2, where $\boldsymbol{X}$ represent vector of concentration of

each species and $\boldsymbol{Y}$ represents the vector with the number of molecules of each species. For example, if a Hill equation has the form $f(x) = (x^n)/(x^n + \theta)$ then its propensity becomes $g(z) = \Omega f(x/\Omega) = \Omega((x/\Omega)^n)/((x/\Omega)^n + \theta)$. The degradation rates do not change since if $f(x) = kx$ then $\Omega f(x/\Omega) = \Omega kx/\Omega = kx$.

On the other hand, the stoichiometry vectors are given in Table 5.4 and the stoichiometry matrix $\boldsymbol{S}$ is the matrix where its columns are the stoichiometry vectors, i.e. $\boldsymbol{S} = (\boldsymbol{\nu}_1, \ldots, \boldsymbol{\nu}_{20})$.

5.6.1 Stochastic simulations

In order to simulate the sequence of reactions, we use the chemical Langevin equation SDE approximation introduced in Section 2.2. From Equation 2.24 and using an Euler-Maruyama method (see Section A.1 in Appendix A), the state of the system at time $t + dt$ is given by the equation:

$$\boldsymbol{Y}(t + dt) \sim \boldsymbol{Y}(t) + \mathcal{N}\left(\boldsymbol{S}\boldsymbol{\omega}(\boldsymbol{Y}(t), \Theta, \Omega)dt, \boldsymbol{S}\text{diag}\left\{\boldsymbol{\omega}(\boldsymbol{Y}(t), \Theta, \Omega)\right\}\boldsymbol{S}^t dt\right) \tag{5.2}$$

We simulate $N = 10,000$ stochastic trajectories representing the changes of the state of N differentiating epiblast cells. As in the deterministic system, we simulate these trajectories from $t_0 = 0$ to $t_1 = 20$ with a time step $dt = 0.01$, which we consider small enough. The parameters take the values in Tables 5.1 and 5.2 and we choose $\Omega = 700$.

As initial conditions, the initial response of the cells to the input signals, W_R and RA_R, and the initial mRNA levels in the N cells are set to 0, while the initial protein levels are sampled from a a normal distribution with mean

$$(y_T, y_S, y_C, y_B) = (0.001, 0.3, 0.001, 0.001)$$

and covariance matrix obtained from the experimental data from Data Set 1 at day 2, normalised to [0,1] by summing the minimum and dividing by the maximum of the whole data set 1. This approach could be improved since the minimum and maximum values of a sample are highly stochastic, this is also why we cannot normalise the data this way as mentioned in Subsection 5.3.1. However, we decided to take this as a first approach. This will be further discussed in Section 5.7.

Each experimental condition is then simulated as in the deterministic case. We divide the interval $[t0, t1] = [0, 20]$ into four subintervals, the extremes of which will be mapped to the days. This way $t = 0$ corresponds to day 2, $t = 5$ corresponds to day 2.5, $t = 10$ corresponds to day 3, $t = 15$ corresponds to day 3.5 and $t = 20$

Table 5.3: *Stochastic model for the proposed gene regulatory network controlling neural and mesoderm specification during vertebrate trunk development.*

Reaction number	Reaction step	Reaction propensities
1	$W \xrightarrow{\alpha_W} W_R$	$\omega_1 = (\alpha_W \times \Omega)\, W$
2	$W_R \xrightarrow{\alpha_W} \varnothing$	$\omega_2 = \alpha_W W_R$
3	$RA \xrightarrow{\alpha_R} RA_R$	$\omega_3 = (\alpha_R \times \Omega)\, RA$
4	$RA_R \xrightarrow{\alpha_W} \varnothing$	$\omega_4 = \alpha_R RA_R$
5	$\varnothing \xrightarrow{v_B} x_T$	$\omega_5 = (v_T \times \Omega) \left[\frac{(\theta_{BT} y_B \times \Omega^{-1})^{n_{BT}} + (\theta_{TT} y_T \times \Omega^{-1})^{n_{TT}}}{1 + (\theta_{BT} y_B \times \Omega^{-1})^{n_{BT}} + (\theta_{TT} y_T \times \Omega^{-1})^{n_{TT}}}\right] \left[\frac{1}{1 + (\theta_{ST} y_S \times \Omega^{-1})^{n_{ST}}}\right] \left[\frac{1}{1 + (\theta_{CT} y_C \times \Omega^{-1})^{n_{CT}}}\right]$
6	$x_T \xrightarrow{k_B} \varnothing$	$\omega_6 = k_T x_T$
7	$x_T \xrightarrow{\alpha_T} y_T$	$\omega_7 = \alpha_T x_T$
8	$y_T \xrightarrow{\alpha_T} \varnothing$	$\omega_8 = \alpha_T y_T$
9	$\varnothing \xrightarrow{v_S} x_S$	$\omega_9 = (v_S \times \Omega) \left[\frac{(\theta_{RS} RA_R \times \Omega^{-1})^{n_{RS}} + (\theta_{SS} y_S \times \Omega^{-1})^{n_{SS}}}{1 + (\theta_{RS} RA_R \times \Omega^{-1})^{n_{RS}} + (\theta_{SS} y_S \times \Omega^{-1})^{n_{SS}}}\right] \left[\frac{1}{1 + (\theta_{WS} W_R \times \Omega^{-1})^{n_{WS}}}\right] \left[\frac{1}{1 + (\theta_{TS} y_T \times \Omega^{-1})^{n_{TS}}}\right]$
10	$x_S \xrightarrow{k_S} \varnothing$	$\omega_{10} = k_S x_S$
11	$x_S \xrightarrow{\alpha_S} y_S$	$\omega_{11} = \alpha_S x_S$
12	$y_S \xrightarrow{\alpha_S} \varnothing$	$\omega_{12} = \alpha_S y_S$
13	$\varnothing \xrightarrow{v_C} x_C$	$\omega_{13} = (v_C \times \Omega) \left[\frac{(\theta_{WC} W_R \times \Omega^{-1})^{n_{WC}}}{1 + (\theta_{WC} W_R \times \Omega^{-1})^{n_{WC}}}\right] \left[\frac{(\theta_{SC} y_S \times \Omega^{-1})^{n_{SC}}}{1 + (\theta_{SC} y_S \times \Omega^{-1})^{n_{SC}}}\right]$
14	$x_C \xrightarrow{k_C} \varnothing$	$\omega_{14} = k_C x_C$
15	$x_C \xrightarrow{\alpha_C} y_C$	$\omega_{15} = \alpha_C x_C$
16	$y_C \xrightarrow{\alpha_C} \varnothing$	$\omega_{16} = \alpha_C y_C$
17	$\varnothing \xrightarrow{v_B} x_B$	$\omega_{17} = (v_B \times \Omega) \left[\frac{(\theta_{WB} W_R \times \Omega^{-1})^{n_{WB}}}{1 + (\theta_{WB} W_R \times \Omega^{-1})^{n_{WB}}}\right] \left[\frac{1}{1 + (\theta_{TB} y_T \times \Omega^{-1})^{n_{TB}}}\right] \left[\frac{1}{1 + (\theta_{SB} y_S \times \Omega^{-1})^{n_{SB}}}\right]$
18	$x_B \xrightarrow{k_B} \varnothing$	$\omega_{18} = k_B x_B$
19	$x_B \xrightarrow{\alpha_B} y_B$	$\omega_{19} = \alpha_B x_B$
20	$y_B \xrightarrow{\alpha_B} \varnothing$	$\omega_{20} = \alpha_B y_B$

Table 5.4: *Set of reactions and corresponding stoichiometric vector of the stochastic model for sub gene regulatory network 2*

Reaction number	Reaction step	Stoichiometric vector
1	$W \xrightarrow{\alpha_W} W_R$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
2	$W_R \xrightarrow{\alpha_W} \varnothing$	$\nu_2 = [-1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
3	$RA \xrightarrow{\alpha_R} RA_R$	$\nu_3 = [\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
6	$x_T \xrightarrow{k_B} \varnothing$	$\nu_4 = [\ 0, -1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
7	$x_T \xrightarrow{\alpha_T} y_T$	$\nu_5 = [\ 1,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
8	$y_T \xrightarrow{\alpha_T} \varnothing$	$\nu_6 = [\ 1,\ 0, -1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
9	$\varnothing \xrightarrow{v_S} x_S$	$\nu_7 = [\ 1,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
10	$x_S \xrightarrow{k_S} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0, -1,\ 0,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
11	$x_S \xrightarrow{\alpha_S} y_S$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
12	$y_S \xrightarrow{\alpha_S} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0, -1,\ 0,\ 0,\ 0,\ 0]^{\mathrm{t}}$
13	$\varnothing \xrightarrow{v_C} x_C$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0,\ 0]^{\mathrm{t}}$
14	$x_C \xrightarrow{k_C} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0, -1,\ 0,\ 0,\ 0]^{\mathrm{t}}$
15	$x_C \xrightarrow{\alpha_C} y_C$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0,\ 0]^{\mathrm{t}}$
16	$y_C \xrightarrow{\alpha_C} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0, -1,\ 0,\ 0]^{\mathrm{t}}$
17	$\varnothing \xrightarrow{v_B} x_B$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1,\ 0]^{\mathrm{t}}$
18	$x_B \xrightarrow{k_B} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0, -1,\ 0]^{\mathrm{t}}$
19	$x_B \xrightarrow{\alpha_B} y_B$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 1]^{\mathrm{t}}$
20	$y_B \xrightarrow{\alpha_B} \varnothing$	$\nu_1 = [\ 1,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0,\ 0, -1]^{\mathrm{t}}$

corresponds to day 4. To simulate No CHIR condition, we compute the N random walks with parameters $W = 0, RA = 0.09$ at all times. CHIR 2-3 condition is simulated with $W = 0.24, RA = 0.09$ from $t = 0$ to $t = 10$ and $W = 0, RA = 0.09$ from $t = 10$ to $t = 20$. And CHIR condition is simulated by setting $W = 0.24, RA = 0.09$ at all times for the N stochastic trajectories. Our results are shown in Figure 5.34.

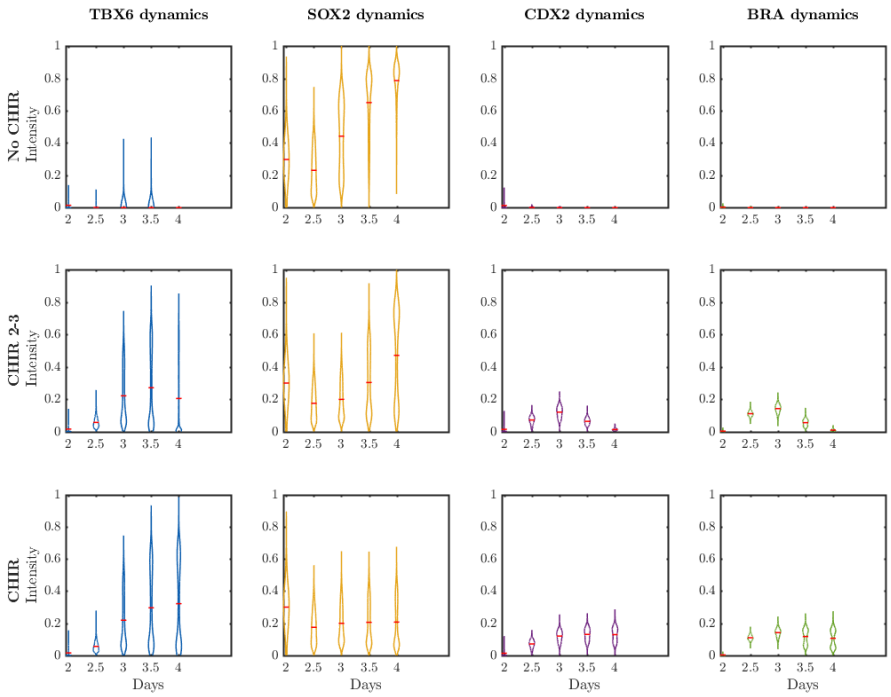


Figure 5.34: *Violin plots of TBX6, SOX2, CDX2 and BRA levels obtained from* $N = 10,000$ *simulated trajectories of the SDE approximation of the proposed stochastic model. First row corresponds to the simulated data under No CHIR condition. Second row of plots corresponds to the CHIR 2-3 simulations. The third row of plots corresponds to the simulated data under CHIR condition. Each column of the plot corresponds to the fitted distribution of the TBX6 (blue), SOX2 (yellow), CDX2 (purple) and BRA (green) simulated mRNA levels. Each violin contains a red line showing the mean of the corresponding distribution.*

Under No CHIR condition, the simulated cells do not express TBX6, CDX2 and BRA at any of the timepoints (the apparent upregulation of TBX6 under CHIR condition is an artefact of the violin plots, only a few simulated cells (4 out of 10,000) upregulate TBX6 and this is due to noise) and they increase their SOX2 levels (see first row of plots in Figures 5.24 and 5.34). The simulated cells become therefore NPs, as observed in the experiments.

The stochastic SOX2 levels under No CHIR condition decrease from day 2

to day 2.5 as in the experimental data, however they increase from day 3 onwards, while it is not until day 3.5 that they increase in the experiments (see first row of plots in Figure 5.24). We will discuss this in Section 5.7.

Under CHIR 2-3 condition, the qualitative behaviour of the simulations is similar to the one observed in the experiments (see second row of plots in Figures 5.24 and 5.34). The simulated cells reach NMP state between day 2.5 and day 3, upregulating CDX2 and BRA levels and downregulating SOX2, which is still present by day 3. We can see, both in the experiments and in the simulated data, that the low SOX2 population starts to upregulate its SOX2 expression at day 3.5, which gradually increases until day 4 when the rest of genes are not expressed. They differentiate into NPs that have gone through NMP state between days 2.5 and 3, as in the experiments. Even though the variance of the simulated CDX2 and BRA expression is lower than in the experimental data, the mean behaviour is the same, except for CDX2. We can notice that the maximal expression times of TBX6 and CDX2 are different between the simulations and the experiments. Also SOX2 is not so much downregulated in by day 3 and it increases faster in our simulations than in the experimental data. We will discuss this in Section 5.7.

Finally, under CHIR condition where Wnt is present from day 2 to day 4, the simulated cells upregulate TBX6 expression from day 2 to day 3, becoming maximal at day 4, as in the experiments. SOX2 is downregulated from day 2 and maintains its expression in a relatively low level for all days both in the experiments and in the simulations. Moreover, by day 3.5 it can be noticed both in CDX2 and BRA expressions that the population of cells divides into two subpopulations, i.e. we can reproduce the bistability observed in the data. Even though the variance of the simulated CDX2 and BRA expression is lower than in the experimental data, the mean behaviour is the same. As with the experimental data under CHIR condition, we also use k-means clustering to untangle the dynamics of the cells that become MPs from the dynamics of the cells that stay as NMPs in the simulated data. Figure 5.35 shows the results of this analysis.

Figure 5.35 shows that indeed the simulate data divides into two populations from day 3 onwards, one corresponding to the NMP population of $SOX2^+$, $CDX2^+$ and BRA^+ cells, and another one corresponding to the MP population of $TBX6^+$ cells, as in the experimental data (see Figure 5.27).

We can see that the stochastic model gives an acceptable qualitative result, reproducing the profiles of the types of cells that are found in experiments under three different experimental conditions. However further work needs to be done to assess some behaviours that the model cannot reproduce at the moment and to also

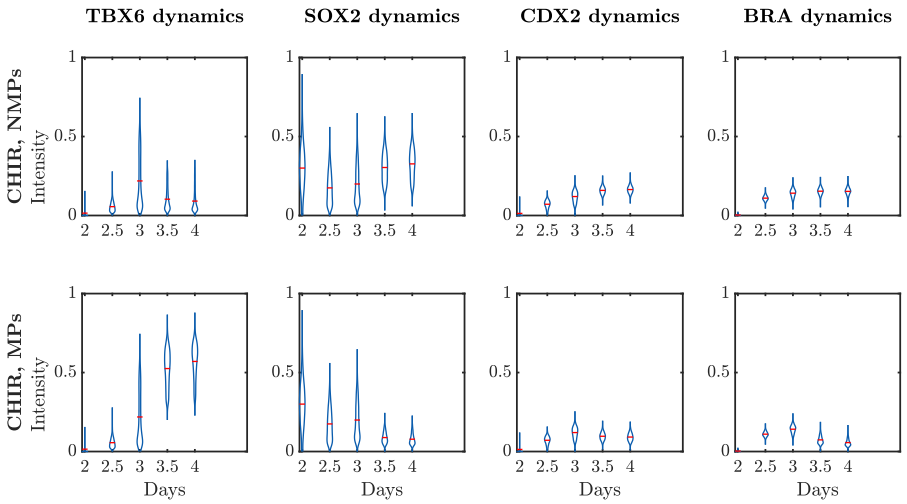


Figure 5.35: *Violin plots of TBX6, SOX2, CDX2 and BRA simulated expression distributions for cells that stay as NMPs (first row) and cells that become MPs (second row) under the simulated CHIR condition. They are obtained by performing k-means clustering after applying the GIGLE method in samples from the simulations at days 3.5 and 4. Each violin contains a red line showing the mean of the sample at the corresponding day and for the corresponding gene.*

quantitatively fit the experimental data. We will discuss these future steps in the following section.

5.7 Discussion

We have investigated the gene regulatory network that controls neural and mesoderm specification of mESCs. In order to do so, we have developed new techniques to process and analyse flow cytometry data.

We have seen that some precautions must be taken to handle flow cytometry data, since it is very sensitive to the efficiency of the staining or different machine set ups. The processing method explained in Subsection 5.3.1 helps to fix some of these errors in the data. Even though this method could be considered somehow subjective, since it relies on the knowledge about the biological process obtained from more accurate techniques such as single-cell qPCR, the majority of the results obtained from it are in agreement with previous findings. Therefore we believe that the method should be trusted if carefully used.

On the other hand, we have not been able to normalise the data sets so that they can be compared. This step is necessary to, for example, check the reproducibility of the experimental results. As we mentioned in Subsection 5.3.1, we tried to scale the data sets to the [0,1] interval by summing the minimum values and

dividing by the maximum values, however this method did not provide good results as these values are highly stochastic. A better method would be to scale the data sets by using the mean values or variances, since they are more robust, however we have not yet found a good method for it.

GIGLE plots have proved to be a very simple but useful visualisation method to reverse engineer the GRN. They show very clearly the dynamics of the gene expression data and the differentiation of stem cells into different cell types. We are interested in testing GIGLE by applying it in other biological settings as well as in gene expression data obtained with other techniques such as single-cell qPCR.

GIGLE plots also show the stochasticity of the system. As we mentioned in Figure 5.7, they show that cells *sit* on the unstable branch of a bifurcation set. This is very interesting and has to do with the stochasticity of the process. The idea of **critical state transitions** in cell differentiation provides an explanation. During the process of differentiation, cells go through different states (or metastates) in which their gene expression profile are stable. Huang et al. (2007), Mojtahedi et al. (2016) propose that the differentiation signal (WNT and RA in our case) may cause destabilization of the high-dimensional gene expression attractor state (which can be considered the epiblast or NMP states in our case) which, in turn, facilitates the (noise-driven) exit from the progenitor attractor to entry into the basin of a new attractor state (NP or MP). This destabilisation can be imagined by a flattening of the potential well which, if it is tilted in a certain direction, would allow the differentiation signal to control the fate decision toward a given fate. This would explain how the cells condense into the NMP state at day 3 under CHIR condition to subsequently open and divide into two subpopulations with some *undecided* cells in the middle, which fall into the corresponding basins of attraction later on (see Figure 5.26).

We have also formulated a minimal gene regulatory model that accounts for the experimental observations of mESCs differentiation into NPs, NMPs and MPs under different Wnt signalling regimes from day 2 to day 4. In contrast to the GRN proposed by Gouti et al. (2017), our analysis shows that BRA does not necessarily downregulate SOX2 and, instead, SOX2 is directly inhibited by Wnt signalling. Our network also includes the posterior marker CDX2 which was not taken into account in the model proposed by Gouti et al. (2017).

The generation of two types of cells under the CHIR experimental condition revealed the multistability of the system, controlled by Wnt and RA signalling. We have showed that the gene regulatory model that we propose can reproduce such dynamics, having attractors for the different expression profiles: the $SOX2^+$ profile

for NPs, the SOX2$^+$, CDX2$^+$, BRA$^+$ profile for NMPs and the TBX6$^+$ for the MPs. We have shown that, for some parameter values, the deterministic model reproduces the behaviour observed in the data under three different experimental conditions.

Moreover, we have defined a stochastic version of such deterministic model and we have shown that the qualitative behaviour is reproduced. However, work needs to be done in order to find parameter values that allow the model to better reproduce the experimental data. If a succesful normalisation of the data is achieved, the next step would be to develop a parameter fitting approach to properly explore the parameter space. By defining a likelihood or distance function by means of the protein expression distributions at each day, one could take advantage of the parameter estimation approaches discussed in Chapter 2.

We have had problems with the timings of TBX6 and CDX2 expression under CHIR 2-3 condition as well as with the delay of the upregulation of SOX2 under No CHIR condition, which we think is also related to the fact that SOX2 level does not decrease so much by day 3 in the CHIR2-3 and CHIR simulations. We suspect that this could be due to FGF signalling that is added to the culture at day 0 and is removed at day 3. It has been shown that FGF represses neural genes such as *Pax6* (Bertrand et al. 2000), but there is no proved inhibition of *Sox2* by FGF. However, in our data, it can be seen that SOX2 is downregulated from day 2 to day 3 under No CHIR condition (see Figure 5.24). Since we did not have enough experimental evidence for this link and we aimed for a minimal model, we did not consider this regulation in our proposed GRN, but it could be a future area to explore.

We have also observed in Data Set 2 (see Figure 5.16), and it can be noticed in Figures 5.14 and 5.27, that BRA starts to be downregulated at day 4 for NMP cells, showing that they will differentiate into posterior neural cells by day 5. Since we did not have available data for the differentiation of cells under the three experimental conditions for five days nor data for the posterior Hox genes, we could not explore the dynamics to reproduce this behaviour in the system. It would be an interesting future extension of our model once the necessary experimental data are available.

Chapter 6

Discussion and conclusions

In this thesis we have developed new approaches to the modelling of cell differentiation and have illustrated them by applying them in different developmental processes.

6.1 Vulval development: a singularity approach

In Chapter 3 we have shown that catastrophe theory can be a key tool for the development of landscape models for cell differentiation. By taking advantage of the fold and the cusp catastrophes, we have shown that we can generate a model that has the characteristics that we required for the process of vulval development in *C. elegans*. We have shown that by using normal forms for the definition of the model, one has a rich understanding of the bifurcations that can arise, which are also controlled by a very small amount of parameters, three in our case. We have shown how to map the signals into the control parameters by a linear relation. By studying this linear relation, we could recognise two possible types of maps from the signal space to the control space, which is important for the identifiability of the parameters. Finally, by using this analysis, we were able to present two types of quasi-potential models for the vulval development and explained how to carry out numerical simmulations from it so that experimental and simulated data can be compared.

In Chapter 4, we took advantage of the ABC SMC method to estimate the parameters of the model with the available experimental data, bypassing the intractable likelihood problem. This method sequentially approximates the posterior distribution of the parameters given the data by sampling parameter values that give more and more similar results to the experimental data in each step of the algorithm. By defining the constraints imposed in the model by the data, the distance function, the kernel distributions and the sequence of thresholds we were allowed to identify the most important parameters in our model, which were then estimated. We consider that this approach improves the one given by Corson & Siggia (2012, 2017) since it accounts for stochasticity and globally explores the

parameter space, providing a real posterior of the parameters. With this approach, we were able to reproduce the results of seven experiments with both the Type I and Type II models defined in Chapter 3, and we did it with significant accuracy. We were also able to reproduce the result of an eighth experiment with both models, and we showed that in this case the approximated posterior obtained with the ABC SMC algorithm approximated the true posterior. However, we were not able to find parameter values that could reproduce both subsets of experimental results at the same time. We inferred from these results that possibly the model that we defined was too restrictive, so we finalised by giving a conjecture for a possible future model by taking advantage of the compactified elliptic umbilic catastrophe.

We would like conclude the discussion of this part of the thesis pointing out that this can be used as a general approach to find quasi-potential landscape models for stem cell differentiation. The key observation is that, instead of trying to look for the mechanistic genotypic model, one can use normal forms to build a model that reproduces the phenotypic behaviour observed in the data. The advantage of using these normal forms is that, first, the class of such normal forms is much smaller and much more controlled than the class of all models and, second, we have a rich understanding of the normal forms from singularity theory and dynamical systems theory as we showed in this thesis. Moreover, the models derived from normal forms have a well-defined vector of control parameters that is typically small in dimension. This is a very important advantage over more mechanistic cellular models, like reaction networks models that, as we saw in Chapter 5, tend to have a very large amount of parameters which have highly correlated effects. One can then map the signals that the cell receives into the control parameters of the system to be able to fit the model to the available experimental data with advanced statistical methods as we proposed in Chapter 4.

6.2 Vertebrate trunk development: GIGLE and bistability

In Chapter 5 we have studied the gene regulatory network that controls neural and mesoderm specification in mouse embryonic stem cells. This is a less understood process, compared to vulval development in *C. elegans*, so we needed to extract the necessary information from the data in order to develop a model for it. We focused our study in the dynamics of four genes, *Tbx6*, *Sox2*, *Cdx2* and *Bra* of which we had protein expression data available from day 2 to day 4 under three experimental conditions (No CHIR, CHIR 2-3, CHIR). These experimental conditions differred

on the Wnt signalling that the cells are exposed to.

Since the data acquisition introduced some errors in the data our collaborators developed a method that, even though had some limitations, corrected them and allowed us to make use of it. We also created a method, GIGLE, that helped us to make sense of the data. GIGLE can be used in gene or protein expression data and shows how the expression of a gene (or a protein) affects the expression of other genes (or proteins). We illustrated this by applying it to the available data. The plots generated clearly showed the critical state transition that the mESCs undergo under CHIR condition, unveiling the bistable nature of the process. Moreover, we were able to disantangle the dynamics of the two types of cells that where generated under the same condition which allowed us to understand the dynamics behind the process.

From this analysis, we were able to identify the different types of cells that the mESCs differentiate to under the three different experimental conditions. If Wnt signalling was not present, cells differentiated into neural progenitor cells, expressing just SOX2. If Wnt signalling was present from day 2 to day 3 and then removed, cells differentiated into neural progenitor cells but going through an intermediate state in which they co-expressed SOX2, CDX2 and BRA, the NMP profile. Finally, if they were constantly exposed to Wnt signalling, the bistability was revealed since two populations of cells arose: some of the cells in the NMP state at day 3 differentiated into mesoderm progenitor cells, expressing just TBX6 by day 4.

We constructed a minimal gene regulatory network that reproduced the bistability observed in the data. Wnt and RA were considered the signal parameters of the system and, under certain values of those parameters, the gene regulatory model had attractors for the different expression profiles: the $SOX2^+$ profile for NPs, the $SOX2^+$, $CDX2^+$, BRA^+ profile for NMPs and the $TBX6^+$ for the MPs. We showed that, for some parameter values, the deterministic model reproduced the behaviour seen in the data under three different experimental conditions. We also produced a stochastic version of it which qualitatively reproduced the behaviour observed in the experimental data.

Interesting future paths can be taken from this point. The gene regulatory network that we have proposed could be extended in order to distinguish between anterior neural progenitors (cells that express SOX2 and SOX1 at day 4) and posterior neural progenitors (cells that upregulate Hox genes by upregulating CDX2 at some point of their development). Since we didn't have this data available, we could not consider this distinction. Moreover, we suspect that there is some relationship between the the FGF signal present in the culture from day 0 to day 3 and the

delay of expression of SOX2 under No CHIR condition. However, this would need further research. Also, the development of a normalisation technique for the experimental data could allow to quantitatively fit the gene regulatory network model to the data. Finally, a very interesting next step would be to develop a phenotypic quasi-potential landscape model that reproduces the dynamics described here by using the techniques explained in Chapters 3–4, and compare it to a genotypic quasi-potential landscape model as recently conceived in Zhou et al. (2012), Guo et al. (2017), Brackston et al. (2018). If one could obtain a genotype-phenotype map (as described in Jaeger & Monk (2015)), that sent the mechanistic GRN variables and parameters into the variables and parameters of a normal form, one could be able to mathematically join Waddington's and Thom's ideas developed many years ago.

Appendix A

Methods

A.1 Euler-Maruyama Method

Consider the n-dimensional Itô process $\boldsymbol{Y} = \{\boldsymbol{Y}_t, t_0 \leq t \leq T\}$ that is governed by the stochastic differential equation (SDE) of the form

$$d\boldsymbol{Y}_t = \boldsymbol{F}(\boldsymbol{Y}_t, \boldsymbol{\theta})dt + \boldsymbol{G}(\boldsymbol{Y}_t, \boldsymbol{\theta})d\boldsymbol{W}_t, \tag{A.1}$$

where $\boldsymbol{\theta}$ represents the parameter vector and $\boldsymbol{W}_t$ is an n-dimensional Wiener process (or, in other words, a Brownian motion (Papoulis & Pillai 2002)). One can interpret the SDE in Equation A.1 as the limit of the difference equation

$$\Delta\boldsymbol{Y}_t = \boldsymbol{F}(\boldsymbol{Y}_t, \boldsymbol{\theta})\Delta t + \boldsymbol{G}(\boldsymbol{Y}, \boldsymbol{\theta})\Delta\boldsymbol{W}_t, \tag{A.2}$$

as Δt gets infinitelly small. For finite Δt, this is known as the Euler-Maruyama approximation of the SDE. Given a discretisation of the time interval $[t_0, T]$,$t_0 < t_0 + \Delta t < t_0 + 2\Delta t < \cdots < t_0 + N\Delta t = T$, i.e. $\Delta t = \frac{T - t_0}{N}$, the **Euler-Maruyama** scheme (Higham 2001, Kloeden & Platen 1992, Wilkinson 2011) is an iterative method that allows to obtain an approximation of a path realisation from $\boldsymbol{X}$ at the points of the discretisation.

The Euler-Maruyama scheme is the simplest strong Taylor approximation, containing only the time and Wiener integrals of multiplicity one from the Itô-Taylor expansion, and usually has order of strong convergence 0.5 (Kloeden & Platen 1992). Even though it is the simplest, its performance is similar to more complicated methods (Higham 2001), therefore we choose it as the method to implement.

In Chapter 2 we introduced the chemical Langevin equation (Equation 2.24), which will be solved in Chapter 5 for a particular n-dimensional system (Section 5.6). The CLE has the form

$$d\boldsymbol{Y}_t = \boldsymbol{S}\boldsymbol{\omega}(\boldsymbol{Y}_t)dt + \sqrt{\boldsymbol{S}\mathrm{diag}\left\{\boldsymbol{\omega}(\boldsymbol{Y}_t)\right\}\boldsymbol{S}^t}d\boldsymbol{W}_t. \tag{A.3}$$

where $\boldsymbol{Y}$ is an n-dimensional Itô process representing the change of n molecular

species in time. If one would like to sample a path from $\boldsymbol{Y}$ satisfying Equation A.3 in the time interval $[t_0, T]$, by choosing a discretisation with time step Δt small, and an initial condition $\boldsymbol{Y}_0$, the iterative Euler-Maruyama method gives the solution at time $t + \Delta t$ from the solution at time t by the equation

$$\boldsymbol{Y}(t+\Delta t) \sim \boldsymbol{Y}(t)+\boldsymbol{S}\boldsymbol{\omega}(\boldsymbol{Y}(t),\Theta,\Omega)\Delta t+\sqrt{\boldsymbol{S}\mathrm{diag}\left\{\boldsymbol{\omega}(\boldsymbol{Y}(t),\Theta,\Omega)\right\}\boldsymbol{S}^t}(\boldsymbol{W}(t+\Delta t)-\boldsymbol{W}(t)) \tag{A.4}$$

where $\boldsymbol{W}(t + \Delta t) - \boldsymbol{W}(t) \sim \mathcal{N}(0, I\Delta t)$ and I is the n-dimensional identity matrix.

In Chapter 3, we make use of this method to solve the 6-dimensional system of stochastic differential equations in Equation 3.36, which can be simplified as

$$d\boldsymbol{X}_t = \boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\theta})dt + \sqrt{2\sigma_{dif}}d\boldsymbol{W}_t, \tag{A.5}$$

where $\boldsymbol{W}_t$ is a white noise.

In this case, using Euler-Maruyama for a discretisation of the interval $[0, 1]$ with time step $\Delta t = 0.02$ and an initial condition $\boldsymbol{X}(0) = \boldsymbol{X}_0$, the evaluation of the path at time $t + dt$ is given by the equation

$$\boldsymbol{X}(t + \Delta t) \sim \boldsymbol{X}(t) + \boldsymbol{F}(\boldsymbol{X}(t), \boldsymbol{\theta})\Delta t + \sqrt{2\sigma_{dif}}(\boldsymbol{W}(t + \Delta t) - \boldsymbol{W}(t)) \tag{A.6}$$

where $\boldsymbol{W}(t + \Delta t) - \boldsymbol{W}(t) \sim \mathcal{N}(0, I\Delta t)$ and I is the 6-dimensional identity matrix.

A.2 Cardano's formula

Cardano's formula (Spiegel 1968) gives the roots of the cubic equation:

$$a_1x^3 + a_2x^2 + a_3x + a_4 = 0 \tag{A.7}$$

where $a_1, a_2, a_3, a_4 \in \mathbb{R}$.

The number of real roots depends on the value of the discriminant $D = Q^3 + R^2$ where

$$Q = \frac{3a_1a_3 - a_2^2}{9a_1^2}, \quad R = \frac{9a_1a_2a_3 - 27a_1^2a_4 - 2a_2^3}{54a_1^3}. \tag{A.8}$$

If $D > 0$, then there is only one real root and the other two are complex conjugates. If $D = 0$, then all the roots are real and at least two are equal. If $D < 0$, then all

the roots are real and unequal. The solutions of Equation A.7 are

$$x_1 = S + T - \frac{a_2}{3a_1} \tag{A.9}$$

$$x_2 = -\frac{S+T}{2} - \frac{a_2}{3a_1} + \frac{i\sqrt{3}}{2}(S-T) \tag{A.10}$$

$$x_3 = -\frac{S+T}{2} - \frac{a_2}{3a_1} - \frac{i\sqrt{3}}{2}(S-T) \tag{A.11}$$

where

$$S = \sqrt[3]{R+\sqrt{D}}, \quad T = \sqrt[3]{R-\sqrt{D}}.$$

In our case, the cubic equation comes from the cusp and it has the form

$$-4x^3 - 2ax - b = 0, \tag{A.12}$$

i.e. $a_1 = -4$, $a_2 = 0$, $a_3 = a$ and $a_4 = b$.

The discriminant D in this case is equal to

$$D = Q^3 + R^2 = \frac{a^3}{2^3 3^3} + \frac{b^2}{2^6} = \frac{8a^3 + 27b^2}{1728}. \tag{A.13}$$

Therefore, the number of real roots depends on the sign of $8a^3 + 27b^2$, which we called $\Delta(a,b) = \Delta = 8a^3 + 27b^2$ in Section 2.3 and Subsection 3.3.2, and is the equation of the bifurcation set of the cusp. So, as expected, the number of roots or equilibria are controlled by the bifurcation set.

Also, in this case, we can write S and T as

$$S = \frac{\sqrt[3]{-b+\sqrt{\Delta/27}}}{2}, \quad T = \frac{\sqrt[3]{-b-\sqrt{\Delta/27}}}{2} \tag{A.14}$$

Considering S and T in Equation A.14 and since $a_2 = 0$, the roots of the cubic equation in A.12 are then given by:

1. If $\Delta > 0$, there will be just one real root given by

$$x = S + T$$

2. If $\Delta = 0$, then $S = T$ and there will be two real roots given by

$$x_1 = 2S = -\sqrt[3]{b}$$
$$x_2 = -S = \frac{\sqrt[3]{b}}{2}$$

3. If $\Delta < 0$, then there will be three real roots given by

$$x_1 = S + T$$
$$x_2 = -\frac{S+T}{2} + \frac{i\sqrt{3}}{2}(S-T)$$
$$x_3 = -\frac{S+T}{2} - \frac{i\sqrt{3}}{2}(S-T)$$

A.3 SDE approximation based on LNA

We are interested in obtaining an approximation of the solution of the system of SDEs given by:

$$\begin{cases} \frac{dx}{dt} &= \frac{1}{\tau}[\chi(y)(-4x^3 - 2a_0x - b_0) - (1-\chi(y))x] + \eta_{01}(t) \\ \frac{dy}{dt} &= \frac{1}{\tau}[y((y-M)^2 + c_0)] + \eta_{02}(t) \end{cases} \tag{A.15}$$

where a_0, b_0, c_0 are obtained from Equation (3.38) with $EGF_0 = 0$ and $NOTCH_0 = 0$ and η_{01} and η_{02} represent independent white noises with variance $2\sigma_{dif}$.

Let us rewrite equation (A.15) the following way:

$$\frac{d\boldsymbol{X}}{dt} = \boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\theta}_0) + \vec{\boldsymbol{\eta}}_0(t), \tag{A.16}$$

where $\boldsymbol{X} = (x, y)^T$, $\vec{\boldsymbol{\eta}}_0(t) = (\eta_{01}(t), \eta_{02}(t))^T$, $\boldsymbol{\theta}_0 = (a_0, b_0, c_0)$ and

$$\boldsymbol{F}(\boldsymbol{X}, \boldsymbol{\theta}_0) = \begin{pmatrix} \frac{1}{\tau}[\chi(y)(-4x^3 - 2a_0x - b_0) - (1-\chi(y))x] \\ \frac{1}{\tau}[y((y-M)^2 + c_0)], \end{pmatrix}$$

with initial condition centered around $\boldsymbol{X}(0) = (0, M)^T$.

Furthermore,

$$\langle \boldsymbol{\eta}_{0i}(t)\boldsymbol{\eta}_{0j}(t')\rangle = 2\sigma_{dif}\delta_{ij}\delta(t-t').$$

This means that $\vec{\boldsymbol{\eta}}_0(t)$ is a Gaussian random noise in which $\langle \boldsymbol{\eta}_{0i}(t)\boldsymbol{\eta}_{0i}(t')\rangle = 2\sigma_{dif}\delta(t-t')$ (i.e. the value $\boldsymbol{\eta}_{0i}(t)$ for any time t is a random variable that is statistically in-

dependent of its entire history before t) so each $\boldsymbol{\eta}_{0i}(t)$ is a **white noise**. And also $\langle \boldsymbol{\eta}_{0i}(t)\boldsymbol{\eta}_{0j}(t')\rangle = 2\sigma_{dif}\delta_{ij}\delta(t-t')$, so the components of $\vec{\boldsymbol{\eta}}_0$ are independent from each other. Taking all these details into account, we can say that, fixed t, $\vec{\boldsymbol{\eta}}_0(t)$ follows a multivariate Gaussian distribution:

$$\vec{\boldsymbol{\eta}}_0(t) \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2\sigma_{dif} & 0 \\ 0 & 2\sigma_{dif} \end{pmatrix}\right) = \mathcal{N}(0, S^2). \tag{A.17}$$

We can now write equation (A.3) as:

$$d\boldsymbol{X}(t) = \boldsymbol{F}(\boldsymbol{X}(t), \boldsymbol{\theta}_0)dt + \boldsymbol{S}\vec{\boldsymbol{\xi}}_0(t)dt, \tag{A.18}$$

where $\vec{\boldsymbol{\xi}}_0(t)$ is a Gaussian white noise with zero mean and identity covariance matrix. We can write $\vec{\boldsymbol{\xi}}_0(t) = d\boldsymbol{W}(t)/dt$ where $\boldsymbol{W}(t)$ is a Brownian motion or Wiener process. This way we can rewrite our SDE as

$$d\boldsymbol{X}_t = \boldsymbol{F}(\boldsymbol{X}_t, \boldsymbol{\theta}_0)dt + \boldsymbol{S}d\boldsymbol{W}_t, \tag{A.19}$$

with initial condition around $\boldsymbol{X}(0) = (0, M)^T$.

We are going to derive an approximation of the SDE in Equation A.15 based on a similar method to the one in (Fearnhead et al. 2014, Wallace 2010) for deriving the Linear Noise Approximation (LNA). Here, we use this method in our more simple case, where the diffusion matrix is constant. Our goal is to approximate the stationary distribution in the basin of attraction of the 3° fate. The idea is to split the SDE in into a deterministic path, $\boldsymbol{x}(t)$, and a stochastic perturbation from this path, assuming that the perturbation is "small" relative to the deterministic solution. We can assume that the perturbation is small because we are taking a small noise and the landscape is fixed, so the trajectory of the system will stay in the same basin of attraction. To express that the perturbation is "small" and to aid in the collection of terms of similar size, we will rewrite $\boldsymbol{S}$ as $\varepsilon\boldsymbol{S}$. Thus, solving (A.19) is the same as solving:

$$\begin{cases} d\boldsymbol{X}(t) = \overbrace{\boldsymbol{F}(\boldsymbol{X})dt}^{\text{Deterministic part}} + \overbrace{\varepsilon\boldsymbol{S}d\boldsymbol{W}(t)}^{\text{Perturbation}} \\ \boldsymbol{X}(0) = (0, M)^T \end{cases} \tag{A.20}$$

The **deterministic solution** is obtained by solving the ODE system:

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt} = \boldsymbol{F}(\boldsymbol{x}(t)) \\ \boldsymbol{x}(0) = \boldsymbol{x}_0 \end{cases} \tag{A.21}$$

The point $\boldsymbol{x}(T)$ will give us the solution of the system in equation (A.20) at time $t = T$ if there was zero noise.

As previously said, we assume that the perturbation is small of order ε, i.e. $\|\boldsymbol{X} - \boldsymbol{x}\|$ is $O(\varepsilon)$. Setting $\boldsymbol{\mathcal{M}}(t) = (\boldsymbol{X}(t) - \boldsymbol{x}(t))/\varepsilon$ and taking a Taylor expansion of $\boldsymbol{F}$ around $\boldsymbol{x}(t)$, we get:

$$\boldsymbol{F}(\boldsymbol{X}(t)) = \boldsymbol{F}(\boldsymbol{x}(t)) + \frac{\partial \boldsymbol{F}_i}{\partial x_j}(\boldsymbol{x}(t))(\boldsymbol{X}(t) - \boldsymbol{x}(t)) + O(\varepsilon^2), \tag{A.22}$$

thus

$$\varepsilon d\boldsymbol{\mathcal{M}}(t) = d(\boldsymbol{X}(t) - \boldsymbol{x}(t)) = d\boldsymbol{X}(t) - d\boldsymbol{x}(t) =$$

$$= \boldsymbol{F}(\boldsymbol{x}(t))dt + \tfrac{\partial \boldsymbol{F}_i}{\partial x_j}(\boldsymbol{x}(t))(\boldsymbol{X}(t) - \boldsymbol{x}(t))dt + O(\varepsilon^2)dt + \varepsilon \boldsymbol{S} d\boldsymbol{W}(t) - \boldsymbol{F}(\boldsymbol{x}(t))dt$$

$$\approx \tfrac{\partial \boldsymbol{F}_i}{\partial x_j}(\varphi(t))\varepsilon \boldsymbol{\mathcal{M}}(t)dt + \varepsilon \boldsymbol{S} d\boldsymbol{W}(t),$$

and we can make the approximation

$$d\boldsymbol{\mathcal{M}}(t) = \boldsymbol{J}(t)\boldsymbol{\mathcal{M}}(t)dt + \boldsymbol{S}d\boldsymbol{W}(t) \tag{A.23}$$

where $\boldsymbol{J}(t) = \dfrac{\partial \boldsymbol{F}_i}{\partial x_j}(\boldsymbol{x}(t))$ is the Jacobian matrix of $\boldsymbol{F}$ and the initial condition is $\boldsymbol{\mathcal{M}}(0) = \boldsymbol{\mathcal{M}}_0 = (\boldsymbol{X}(0) - \boldsymbol{x}_0)/\varepsilon$.

The increment in (A.23) is a linear combination of Gaussians, so $\boldsymbol{\mathcal{M}}(t)$ has a Gaussian distribution for all t. In particular, $\boldsymbol{\mathcal{M}}(T)$ will be the distribution of the error when the solution of the SDE is approximated by the deterministic solution.

If we define $\boldsymbol{\mu}(t) = \mathbb{E}[\boldsymbol{\mathcal{M}}(t)]$ and $\boldsymbol{\Phi}(t) = \mathrm{Var}[\boldsymbol{\mathcal{M}}(t)]$, by the linearity of the expectation and Eq. A.23, (Fearnhead et al. 2014):

$$\frac{d\boldsymbol{\mu}(t)}{dt} = \mathbb{E}\left[\frac{d\boldsymbol{\mathcal{M}}(t)}{dt}\right] = \mathbb{E}[\boldsymbol{J}(t)\boldsymbol{\mathcal{M}}(t) + \boldsymbol{S}\vec{\boldsymbol{\xi}}(t)] = \mathbb{E}[\boldsymbol{J}(t)\boldsymbol{\mathcal{M}}(t)] + \boldsymbol{S}\mathbb{E}[\vec{\boldsymbol{\xi}}(t)] = \boldsymbol{J}(t)\boldsymbol{\mu}(t),$$

$$\begin{aligned}\frac{\mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t]}{dt} &= \mathbb{E}\left[\tfrac{d\boldsymbol{\mathcal{M}}(t)}{dt}\boldsymbol{\mathcal{M}}(t)^t\right] + \mathbb{E}\left[\boldsymbol{\mathcal{M}}(t)\tfrac{d\boldsymbol{\mathcal{M}}(t)^t}{dt}\right] \\ &= \boldsymbol{J}(t)\mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t] + \mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t]\boldsymbol{J}(t)^t + \boldsymbol{S}\boldsymbol{S}^t.\end{aligned}$$

Hence

$$\frac{d\boldsymbol{\Phi}(t)}{dt} = \frac{\mathrm{Var}[\boldsymbol{\mathcal{M}}(t)]}{dt} = \frac{\mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t]}{dt} - \frac{\boldsymbol{\mu}(t)\boldsymbol{\mu}(t)^t}{dt} =$$

$$= \boldsymbol{J}(t)\mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t] + \mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t]\boldsymbol{J}(t)^t + \boldsymbol{S}\boldsymbol{S}^t - \boldsymbol{J}(t)\mathbb{E}[\boldsymbol{\mathcal{M}}(t)\boldsymbol{\mathcal{M}}(t)^t] - \\ - \boldsymbol{J}(t)\boldsymbol{\mu}(t)\boldsymbol{\mu}(t)^t - \boldsymbol{\mu}(t)\boldsymbol{\mu}(t)^t =$$

$$= \boldsymbol{J}(t)\boldsymbol{\Phi}(t) + \boldsymbol{\Phi}(t)J(t)^t + \boldsymbol{S}\boldsymbol{S}^t$$

From this, we infer that the mean and variance of $\boldsymbol{\mathcal{M}}(t)$ are the solutions of the ODEs:

$$\frac{d\boldsymbol{\mu}(t)}{dt} = \boldsymbol{J}(t)\boldsymbol{\mu}(t) \tag{A.24}$$

$$\frac{d\boldsymbol{\Phi}(t)}{dt} = \boldsymbol{J}(t)\boldsymbol{\Phi}(t) + \boldsymbol{\Phi}(t)J(t)^t + \boldsymbol{S}\boldsymbol{S}^t \tag{A.25}$$

Now, if $\mathbb{E}(\boldsymbol{X}(0)) = (0, M)$, choosing $\boldsymbol{x}(0) = (0, M)$ as initial condition for the deterministic solution we get the initial conditions $\boldsymbol{\mu}(0) = \boldsymbol{X}_0 - \boldsymbol{x}_0 = 0$ (so the solution is $\boldsymbol{\mu}(t) = 0$) and $\boldsymbol{\Phi}(0) = 0$. Computing the solutions for equations (A.21), (A.24) and (A.25), we obtain that the solution $\boldsymbol{X}(t)$ to the first SDE of our model (Eq. (A.3)) follows a Gaussian distribution with mean $\boldsymbol{x}(t)$ and variance $\boldsymbol{\Phi}(t)$ (Fearnhead et al. 2014). In particular, for $t = T$, with T large, $\boldsymbol{X}(T)$ gives us an approximation of the stationary distribution of the system described in Equation (A.15).

Appendix B

Linear transformations from signal to control space

Here we study the possible intersections of the bifurcation set $\mathcal{B}$ of the system in Equation 3.5 given by

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{(a,b,c) \in \mathbb{R}^3 : 8a^3 + 27b^2 = 0\} \cup \{(a,b,0) : a,b \in \mathbb{R}\} \tag{B.1}$$

and the plane π_T defined by the affine transformation T from the signal space to the control space given by

$$\begin{gathered} T : \mathbb{R}^2 \longrightarrow \mathbb{R}^3 \\ X = (EGF, NOTCH) \mapsto T(X) = \begin{pmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \\ m_{31} & m_{32} \end{pmatrix} \begin{pmatrix} EGF \\ NOTCH \end{pmatrix} + \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix}, \end{gathered} \tag{B.2}$$

from which we obtain that

$$\pi_T = \{(a,b,c) \in \mathbb{R}^3 : Aa + Bb + Cc = D\} \tag{B.3}$$

with

$$A = m_{31}m_{22} - m_{21}m_{32} \tag{B.4}$$

$$B = m_{11}m_{32} - m_{31}m_{12} \tag{B.5}$$

$$C = m_{12}m_{21} - m_{11}m_{22} \tag{B.6}$$

$$D = -Aq_1 - Bq_2 - Cq_3. \tag{B.7}$$

We will study the different intersections depending on the values of A, B, C and D. First, we should notice that A, B, C cannot be all equal to zero, otherwise π_T would be a point or $\mathbb{R}^3$. With this in mind, let us study the different cases.

B.1 Case 1: $A = B = 0$, $C \neq 0$

In this case, π_T is the plane $c = D/C$. If $D = 0$, the plane does not intersect the bifurcation set transversally, in fact $\pi_T = \mathcal{B}_1$. The attractor corresponding to fate 3° is always degenerate in this case, therefore we exclude this posibility.

Otherwise, if $D \neq 0$, the plane would not intersect the set $\mathcal{B}_1$ (see Figure B.1), and therefore the attractor corresponding to fate 3° is either not present ($D/C > 0$) or always present in the system ($D/C < 0$).

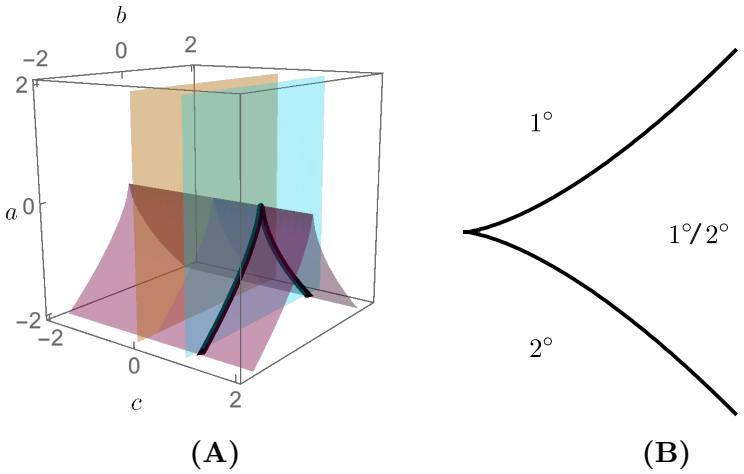
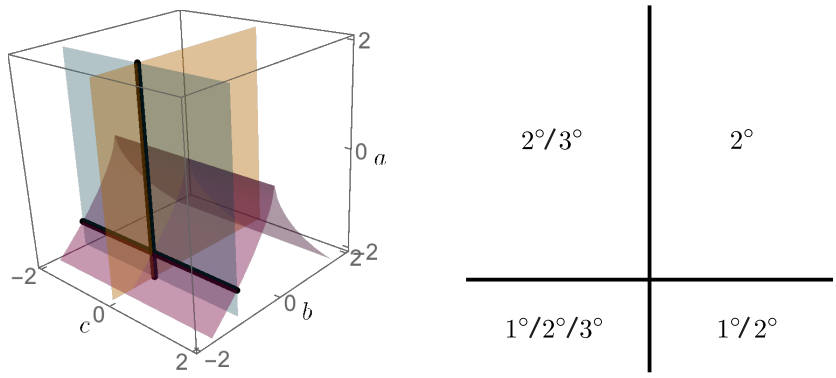


Figure B.1: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $A = B = 0$, $C \neq 0$, $D/C > 0$. I.e. π_T is the plane $c = D/C$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) The different regions determined on the plane π_t by the bifurcation set. In each region, we have specified the fates which attractors would be present in the state space for a parameter value taken in the corresponding region.*

B.2 Case 2: $A = C = 0$, $B \neq 0$

In this case, π_T is the plane $b = D/B$. Figure B.2 shows an example of this configuration. A characteristic of this configuration is that the signal space will not have regions where the system is monostable for both 1° and 2°. I.e, depending on the value on the sign of D/B, the signal space will contain one only monostable region which will correspond to 1° fate if $D/B > 0$, or to 2° if $D/B < 0$ (as in Figure B.2).



(A) **(B)**

Figure B.2: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $A = C = 0$, $B \neq 0$, $D/B < 0$. I.e. π_T is the plane $b = D/B$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) The different regions determined on the plane π_t by the bifurcation set. In each region, we have specified the fates which attractors would be present in the state space for a parameter value taken in the corresponding region.*

B.3 Case 3: $B = C = 0$, $A \neq 0$

In this case, π_T is the plane $a = D/A$. Figure B.3 shows an example of this configuration. In order for the system to have a tristable region in the signal space, D/A should be negative. For D/A the plane is tangent to the cusp line. For $D/A > 0$, the plane π_T does not intersect the cuspoidal cylinder.

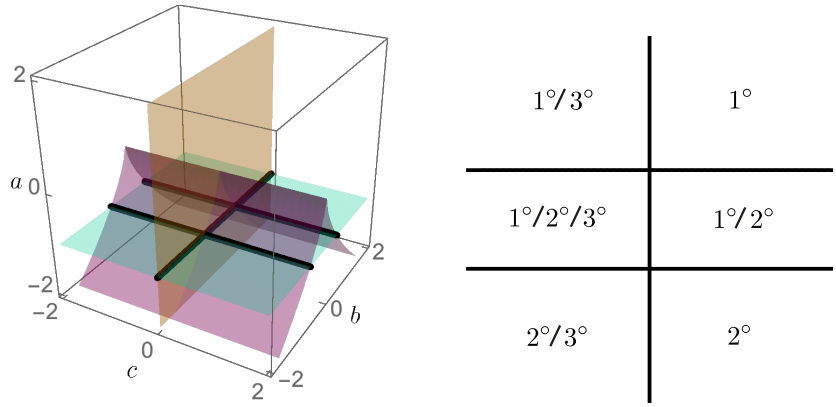
B.4 Case 4: $A = 0$, $B \neq 0 \neq C$

In this case, π_T is the plane $Bb + Cc = D$. Figure B.4 shows an example of this configuration.

The intesecton of π_T and $\mathcal{B}_1$ is the line $c = 0$, $b = D/B$. The intersection of π_T and $\mathcal{B}_2$ is given by the equations

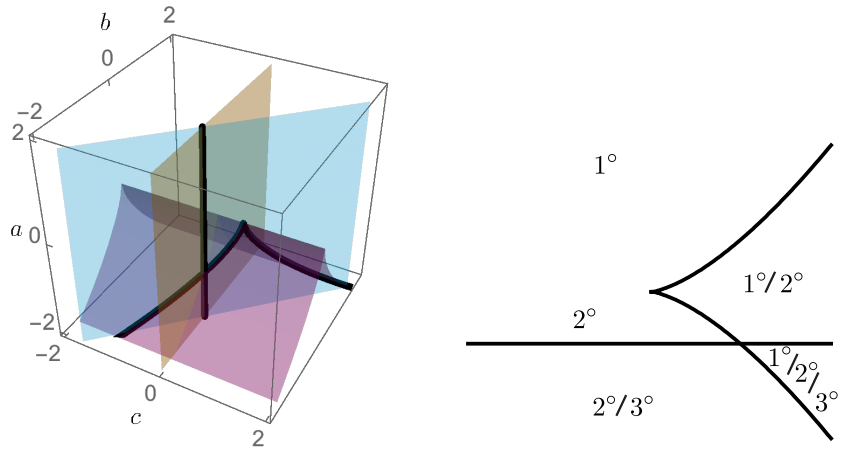
$$\begin{cases} Bb + cC &= D \\ 8a^3 + 27b^2 &= 0, \end{cases}$$

which is a cusp (as we can see in Figure B.4).



(A) **(B)**

Figure B.3: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $B = C = 0$, $A \neq 0$, $D/A < 0$. I.e. π_T is the plane $a = D/A$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) The different regions determined on the plane π_t by the bifurcation set. In each region, we have specified the fates which attractors would be present in the state space for a parameter value taken in the corresponding region.*



(A) **(B)**

Figure B.4: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $A = 0$, $B \neq 0 \neq C$. I.e. π_T is the plane $Bb + Cc = D$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) The different regions determined on the plane π_t by the bifurcation set. In each region, we have specified the fates which attractors would be present in the state space for a parameter value taken in the corresponding region.*

B.5 Case 5: $B = 0$, $A \neq 0 \neq C$

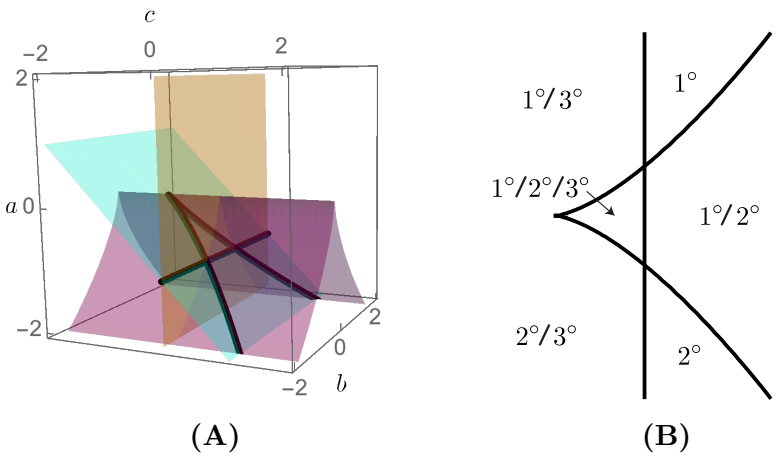
In this case, π_T is the plane $Aa + Cc = D$. Figure B.5 shows an example of this configuration. This is a special case of the one studied in Subsection 3.3.5 where, in

this case, $B = 0$.

The intersection of π_T and $\mathcal{B}_1$ is the line $c = 0$, $a = D/A$. On the other hand, the intersection of π_T and $\mathcal{B}_2$ is a cuspoidal curve given by the equations

$$\begin{cases} Aa + Cc &= D \\ 8a^3 + 27b^2 &= 0. \end{cases}$$

In order for the plane to contain a tristable region, the cusp-like shape and the like must intersect. As explained in Subsection 3.3.5, in order to achieve this we must impose that if $AC > 0$ then $AD < 0$, otherwise if $AC < 0$ then $AD > 0$.



(A) **(B)**

Figure B.5: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $B = 0$, $A \neq 0 \neq C$. I.e. π_T is the plane $Aa + Cc = D$. In this particular case, $AC > 0$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) The different regions determined on the plane π_t by the bifurcation set. In each region, we have specified the fates which attractors would be present in the state space for a parameter value taken in the corresponding region.*

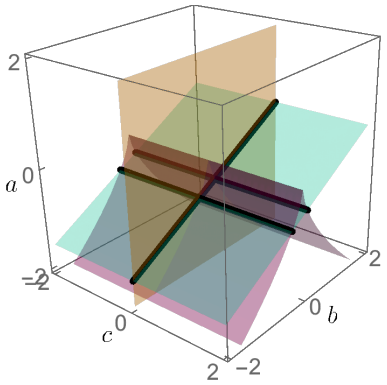
B.6 Case 6: $C = 0$, $A \neq 0 \neq B$

In this case, π_T is the plane $Aa + Bb = D$. Figure B.6 shows an example of this configuration.

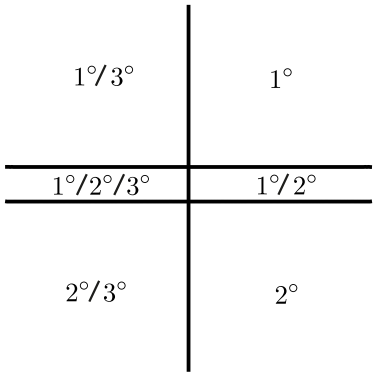
The intersection of π_T and $\mathcal{B}_1$ is the line $c = 0$, $Aa + Bb = D$. On the other hand, the intersection of π_T and $\mathcal{B}_2$ will be on 0, 1, 2 or 3 lines depending on the number of intersections of the lines

$$\begin{cases} Aa + Bb &= D \\ 8a^3 + 27b^2 &= 0. \end{cases}$$

in the $c = 0$ plane. In Figure B.6, the plane π_T and the cuspoidal cylinder $\mathcal{B}_2$ intersect in two lines.



(A)



(B)

Figure B.6: *Example of the intersection $\pi_T \cap \mathcal{B}$ for a T such that $C = D/C$, $D/C > 0$. (A) Intersection of the plane π_T and the bifurcation set $\mathcal{B}$. The set $\mathcal{B}_1$ is plotted in orange, the set $\mathcal{B}_2$ is plotted in purple and the plane π_T is plotted in cyan. The intersection is marked by thick black lines. (B) Attractors present in the signal space for this configuration.*

B.7 Case 7: $A \neq 0, B \neq 0, C \neq 0$

This is the typical case, and contains the the rest of the possible intersections.

BIBLIOGRAPHY

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K. & Walter, P. (2002), *Molecular biology of the cell, 4th edition*, Garland Science.

Alon, U. (2006), *An introduction to systems biology: design principles of biological circuits*, CRC press.

Anderson, D. F. & Kurtz, T. G. (2011), 'Continuous time Markov chain models for chemical reaction networks', *Design and Analysis of Biomolecular Circuits* pp. 3–42.

Baker, J. C., Beddington, R. S. & Harland, R. M. (1999), 'Wnt signaling in Xenopus embryos inhibits bmp4 expression and activates neural development.', *Genes & development* **13**(23), 3149–59.

Beaumont, M. A., Cornuet, J.-M., Marin, J.-M. & Robert, C. P. (2009), 'Adaptive approximate Bayesian computation', *Biometrika* pp. 1–8.

Beaumont, M. A., Zhang, W. & Balding, D. J. (2002), 'Approximate bayesian computation in population genetics', *Genetics* **162**(4), 2025–2035.

Bertrand, N., Médevielle, F. & Pituello, F. (2000), 'Fgf signalling controls the timing of pax6 activation in the neural tube', *Development* **127**(22), 4837–4843.

Boyer, L. A., Tong, I. L., Cole, M. F., Johnstone, S. E., Levine, S. S., Zucker, J. P., Guenther, M. G., Kumar, R. M., Murray, H. L., Jenner, R. G., Gifford, D. K., Melton, D. A., Jaenisch, R. & Young, R. A. (2005), 'Core transcriptional regulatory circuitry in human embryonic stem cells', *Cell* **122**(6), 947–956.

Brackston, R., Lakatos, E. & Stumpf, M. P. (2018), 'Transition State Characteristics During Cell Differentiation', *bioRxiv* p. 264143.

Brooks, S., Gelman, A., Jones, G. & Meng, X.-L. (2011), *Handbook of markov chain monte carlo*, CRC press.

Brown, J. M. & Storey, K. G. (2000), 'A region of the vertebrate neural plate in which neighbouring cells can adopt neural or epidermal fates The caudal neural plate

contains both prospective neural and epidermal cells', *Current Biology* **10**, 869–872.

Christian, P. R. & Casella, G. (1999), 'Monte carlo statistical methods'.

Claverie, J. M. (2001), 'Gene number. What if there are only 30,000 human genes?', *Science* **291**(5507), 1255–7.

Corson, F., Couturier, L., Rouault, H., Mazouni, K. & Schweisguth, F. (2017), 'Self-organized notch dynamics generate stereotyped sensory organ patterns in drosophila', *Science* **356**(6337), eaai7407.

Corson, F. & Siggia, E. D. (2012), 'Geometry, epistasis, and developmental patterning', *Proceedings of the National Academy of Sciences* **109**(15), 5568–5575.

Corson, F. & Siggia, E. D. (2017), 'Gene free methodology for cell fate dynamics during development', *eLife* **6**, e30743.

Craciun, G., Tang, Y. & Feinberg, M. (2006), 'Understanding bistability in complex enzyme-driven reaction networks', *Proceedings of the National Academy of Sciences* **103**(23), 8697–8702.

Davidson, E. H. (2010), *The regulatory genome: gene regulatory networks in development and evolution*, Elsevier.

Fall, C. P. (2002), *Computational cell biology*, Springer, New York.

Fearnhead, P., Giagos, V. & Sherlock, C. (2014), 'Inference for reaction networks using the linear noise approximation', *Biometrics* .

Filippi, S., Barnes, C. P., Cornebise, J. & Stumpf, M. P. (2013), 'On optimality of kernels for approximate Bayesian computation using sequential Monte Carlo', *Statistical Applications in Genetics and Molecular Biology* **12**(1).

Fisher, J., Piterman, N., Hajnal, A. & Henzinger, T. A. (2007), 'Predictive modeling of signaling crosstalk during C. elegans vulval development.', *PLOS Computational Biology* **3**(5), e92.

Fisher, J., Piterman, N., Hubbard, E. J. A., Stern, M. J. & Harel, D. (2005), 'Computational insights into Caenorhabditis elegans vulval development.', *Proceedings of the National Academy of Sciences* **102**(6), 1951–6.

Gillespie, D. T. (1977), 'Exact stochastic simulation of coupled chemical reactions', *The journal of physical chemistry* **81**(25), 2340–2361.

Gillespie, D. T. (2000), 'Chemical Langevin equation', *Journal of Chemical Physics* **113**(1), 297–306.

Giurumescu, C. A., Sternberg, P. W. & Asthagiri, A. R. (2006), 'Intercellular coupling amplifies fate segregation during Caenorhabditis elegans vulval development.', *Proceedings of the National Academy of Sciences* **103**(5), 1331–6.

Gonze, D., Halloy, J., Leloup, J.-C. & Goldbeter, A. (2003), 'Stochastic models for circadian rhythms: effect of molecular noise on periodic and chaotic behaviour Modèles stochastiques pour les rythmes circadiens : effet du bruit moléculaire sur les comportements périodiques et chaotiques', *C. R. Biologies* **326**, 189–203.

Gouti, M., Delile, J., Stamataki, D., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. & Briscoe, J. (2017), 'A Gene Regulatory Network Balances Neural and Mesoderm Specification during Vertebrate Trunk Development', *Developmental Cell* **41**(3), 243–261.e7.

Gouti, M., Metzis, V. & Briscoe, J. (2015), 'The route to spinal cord cell types: A tale of signals and switches', *Trends in Genetics* **31**(6), 282–289.

Gouti, M., Tsakiridis, A., Wymeersch, F. J., Huang, Y., Kleinjung, J., Wilson, V. & Briscoe, J. (2014), 'In Vitro Generation of Neuromesodermal Progenitors Reveals Distinct Roles for Wnt Signalling in the Specification of Spinal Cord and Paraxial Mesoderm Identity', *PLOS Biology* **12**(8).

Greenwald, I. S., Sternberg, P. W. & Horvitz, H. R. (1983), 'The lin-12 locus specifies cell fates in Caenorhabditis elegans.', *Cell* **34**(2), 435–444.

Guckenheimer, J. (1978), 'The catastrophe controversy', *The Mathematical Intelligencer* **1**(1), 15–20.

Guo, J., Lin, F., Zhang, X., Tanavde, V. & Zheng, J. (2017), 'Netland: quantitative modeling and visualization of waddington's epigenetic landscape using probabilistic potential', *Bioinformatics* **33**(10), 1583–1585.

Haario, H., Saksman, E., Tamminen, J. et al. (2001), 'An adaptive metropolis algorithm', *Bernoulli* **7**(2), 223–242.

Hastings, W. K. (1970), 'Monte carlo sampling methods using markov chains and their applications', *Biometrika* **57**(1), 97–109.

Henrique, D., Abranches, E., Verrier, L. & Storey, K. G. (2015), 'Neuromesodermal progenitors and the making of the spinal cord', *Development* **142**(17), 2864–2875.

Higham, D. J. (2001), 'An Algorithmic Introduction to Numerical Simulation of Stochastic Differential Equations', *SIAM Review* **43**(3), 525–546.

Hoyos, E., Kim, K., Milloz, J., Barkoulas, M., Pénigault, J. B., Munro, E. & Félix, M. A. (2011), 'Quantitative variation in autocrine signaling and pathway crosstalk in the Caenorhabditis vulval network', *Current Biology* **21**, 527–538.

Huang, S. (2012), 'The molecular and mathematical basis of Waddington's epigenetic landscape: A framework for post-Darwinian biology?', *BioEssays* **34**(2), 149–157.

Huang, S., Guo, Y.-P., May, G. & Enver, T. (2007), 'Bifurcation dynamics in lineage-commitment in bipotent progenitor cells', *Developmental biology* **305**(2), 695–713.

Huang, S. & Kauffman, S. A. (2012), Complex gene regulatory networks–from structure to biological observables: cell fate determination, *in* 'Computational Complexity', Springer, pp. 527–560.

Jaeger, J. & Monk, N. (2014), 'Bioattractors: dynamical systems theory and the evolution of regulatory processes', *The Journal of Physiology* **592**(11), 2267–2281.

Jaeger, J. & Monk, N. (2015), 'Everything flows: A process perspective on life.', *EMBO reports* **16**(9), 1064–7.

Kloeden, P. E. & Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Springer Berlin Heidelberg, Berlin, Heidelberg.

Komatsu, H., Chao, M. Y., Larkins-Ford, J., Corkins, M. E., Somers, G. A., Tucey, T., Dionne, H. M., White, J. Q., Wani, K., Boxem, M. & Hart, A. C. (2008), 'OSM-11 Facilitates LIN-12 Notch Signaling during Caenorhabditis elegans Vulval Development', *PLOS Biology* **6**(8), e196.

Kondo, S. & Miura, T. (2010), 'Reaction-diffusion model as a framework for understanding biological pattern formation', *Science* **329**(5999), 1616–1620.

Kugler, H., Marelly, R., Appleby, L., Fisher, J., Pnueli, A., Harel, D., Stern, M. J., Hubbard, E. J. A. et al. (2008), 'A scenario-based approach to modeling development: A prototype model of c. elegans vulval fate specification', *Developmental Biology* **323**(1), 1–5.

Levine, M. & Davidson, E. H. (2005), 'Gene regulatory networks for development', *Proceedings of the National Academy of Sciences* **102**(14), 4936–4942.

Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C. P. & H Stumpf, M. P. (2014), 'A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation', *Nature protocols* **9**.

Lints, R. & Hall, D. (2009), 'Reproductive system, egglaying apparatus.', *WormAtlas* .

Lu, R., Markowetz, F., Unwin, R. D., Leek, J. T., Airoldi, E. M., MacArthur, B. D., Lachmann, A., Rozov, R., Ma'ayan, A., Boyer, L. A. et al. (2009), 'Systems-level dynamic analyses of fate change in murine embryonic stem cells', *Nature* **462**(7271), 358.

Marin, J.-M., Pudlo, P., Robert, C. P. & Ryder, R. J. (2012), 'Approximate Bayesian computational methods', *Statistics and Computing* **22**, 1167–1180.

Mckinley, T., Deardon, R. & Cook, A. R. (2009), 'Inference in Epidemic Models without Likelihoods', *The International Journal of Biostatistics* **5**(1).

Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H. & Teller, E. (1953), 'Equation of state calculations by fast computing machines', *The journal of chemical physics* **21**(6), 1087–1092.

Milloz, J., Duveau, F., Nuez, I. & Félix, M. A. (2008), 'Intraspecific evolution of the intercellular signaling network underlying a robust developmental system', *Genes & development* **22**(21), 3064–3075.

Minas, G. & Rand, D. A. (2017), 'Long-time analytic approximation of large stochastic oscillators: Simulation, analysis and inference', *PLOS Computational Biology* **13**(7), e1005676.

Mogilner, A., Wollman, R. & Marshall, W. F. (2006), 'Quantitative modeling in cell biology: what is it good for?', *Developmental Cell* **11**(3), 279–287.

Mojtahedi, M., Skupin, A., Zhou, J., Castaño, I. G., Leong-Quong, R. Y. Y., Chang, H., Trachana, K., Giuliani, A. & Huang, S. (2016), 'Cell Fate Decision as High-Dimensional Critical State Transition', *PLOS Biology* **14**(12), e2000640.

Moral, P. D., Doucet, A., Jasra, A., Carlo, M., Monte, S., Moral, C. P. D., Doucet, A. & Jasra, A. (2012), 'An adaptive sequential Monte Carlo method for approximate Bayesian computation', *Statistics and Computing* **22**, 1009–1020.

Nordström, U., Maier, E., Jessell, T. M. & Edlund, T. (2006), 'An Early Role for Wnt Signaling in Specifying Neural Patterns of Cdx and Hox Gene Expression and Motor Neuron Subtype Identity', *PLOS Biology* **4**(8), e252.

Novák, B. & Tyson, J. J. (2008), 'Design principles of biochemical oscillators', *Nature reviews Molecular cell biology* **9**(12), 981.

Papoulis, A. & Pillai, S. U. (2002), *Probability, random variables, and stochastic processes*, Tata McGraw-Hill Education.

Pennisi, E. (2003), 'Human genome. A low number wins the GeneSweep Pool.', *Science* **300**(5625), 1484.

Perez-Carrasco, R., Barnes, C. P., Schaerli, Y., Isalan, M., Briscoe, J. & Page, K. M. (2017), 'The power of the ac-dc circuit: Operating principles of a simple multi-functional transcriptional network motif', *arXiv preprint arXiv:1708.04593* .

Ping, X. & Tang, C. (2015), 'An Atlas of Network Topologies Reveals Design Principles for Caenorhabditis elegans Vulval Precursor Cell Fate Patterning.', *PLOS ONE* **10**(6), e0131397.

Poston, T. & Stewart, I. (2014), *Catastrophe theory and its applications*, Courier Corporation.

Pritchard, J. K., Seielstad, M. T., Perez-Lezaun, A. & Feldman, M. W. (1999), 'Population growth of human y chromosomes: a study of y chromosome microsatellites.', *Molecular biology and evolution* **16**(12), 1791–1798.

Qiu, P., Simonds, E. F., Bendall, S. C., Gibbs Jr, K. D., Bruggner, R. V., Linderman, M. D., Sachs, K., Nolan, G. P. & Plevritis, S. K. (2011), 'Extracting a cellular hierarchy from high-dimensional cytometry data with spade', *Nature biotechnology* **29**(10), 886.

Ribeiro, A., Zhu, R. & Kauffman, S. A. (2006), 'A general modeling strategy for gene regulatory networks with stochastic dynamics', *Journal of Computational Biology* **13**(9), 1630–1639.

Rubin, D. B. et al. (1984), 'Bayesianly justifiable and relevant frequency calculations for the applied statistician', *The Annals of Statistics* **12**(4), 1151–1172.

Saeys, Y., Gassen, S. V. & Lambrecht, B. N. (2016), 'Computational flow cytometry: helping to make sense of high-dimensional immunology data', *Nature Reviews Immunology* **16**(7), 449–462.

Schmid, T. & Hajnal, A. (2015), 'Signal transduction during C. elegans vulval development: a NeverEnding story.', *Current opinion in genetics & development* **32**, 1–9.

Shaye, D. D. & Greenwald, I. (2002), 'Endocytosis-mediated downregulation of LIN-12/Notch upon Ras activation in Caenorhabditis elegans', *Nature* **420**(6916), 686–690.

Silk, D., Filippi, S. & Stumpf, M. P. H. (2012), 'Optimizing Threshold–Schedules for Approximate Bayesian Computation Sequential Monte Carlo Samplers: Applications to Molecular Systems'.

Sisson, S. A., Fan, Y. & Tanaka, M. M. (2007), 'Sequential Monte Carlo without likelihoods.', *Proceedings of the National Academy of Sciences* **104**(6), 1760–5.

Sisson, S. A., Fan, Y. & Tanaka, M. M. (2009), 'Correction for Sisson et al., Sequential Monte Carlo without likelihoods', *Proceedings of the National Academy of Sciences* **106**(39), 16889–16889.

Southan, C. (2004), 'Has the yo-yo stopped? An assessment of human protein-coding gene number', *PROTEOMICS* **4**(6), 1712–1726.

Spiegel, M. R. (1968), 'Mathematical handbook of formulas and tables'.

Sternberg, P. W. (2005), 'Vulval development', *The C. elegans Research Community, WormBook* .

Sternberg, P. W. & Horvitz, H. (1986), 'Pattern formation during vulval development in C. elegans', *Cell* **44**(5), 761–772.

Sternberg, P. W. & Horvitz, H. (1989), 'The combined action of two intercellular signaling pathways specifies three cell fates during vulval induction in C. elegans', *Cell* **58**(4), 679–693.

Streit, A., Berliner, A. J., Papanayotou, C., Sirulnik, A. & Stern, C. D. (2000), 'Initiation of neural induction by FGF signalling before gastrulation', *Nature* **406**(6791), 74–78.

Takemoto, T., Uchikawa, M., Yoshida, M., Bell, D. M., Lovell-Badge, R., Papaioannou, V. E. & Kondoh, H. (2011), 'Tbx6-dependent Sox2 regulation determines neural or mesodermal fate in axial stem cells', *Nature* **470**(7334), 394–398.

Tanaka, M. M., Francis, A. R., Luciani, F. & Sisson, S. (2006), 'Using approximate bayesian computation to estimate tuberculosis transmission parameters from genotype data', *Genetics* **173**(3), 1511–1520.

Thom, R. (1975), *Structura stability and Morphogenesis*, W.A. Benjamin.

Thomson, M., Liu, S. J., Zou, L. N., Smith, Z., Meissner, A. & Ramanathan, S. (2011), 'Pluripotency factors in embryonic stem cells regulate differentiation into germ layers', *Cell* **145**(6), 875–889.

Toni, T. & Stumpf, M. P. H. (2010), 'Simulation-based model selection for dynamical systems in systems and population biology.', *Bioinformatics* **26**(1), 104–10.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. & Stumpf, M. P. (2009), 'Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems', *Journal of the Royal Society Interface* **6**(31), 187–202.

Turner, D. A., Hayward, P. C., Baillie-Johnson, P., Rue, P., Broome, R., Faunes, F. & Arias, A. M. (2014), 'Wnt/β-catenin and FGF signalling direct the specification and maintenance of a neuromesodermal axial progenitor in ensembles of mouse embryonic stem cells', *Development* **141**(22), 4243–4253.

Tzouanacou, E., Wegener, A., Wymeersch, F. J., Wilson, V. & Nicolas, J. F. (2009), 'Redefining the Progression of Lineage Segregations during Mammalian Embryogenesis by Clonal Analysis', *Developmental Cell* **17**(3), 365–376.

van de Ven, C., Bialecka, M., Neijts, R., Young, T., Rowland, J. E., Stringer, E. J., Van Rooijen, C., Meijlink, F., Nóvoa, A., Freund, J.-N., Mallo, M., Beck, F. & Deschamps, J. (2011), 'Concerted involvement of Cdx/Hox genes and Wnt signaling in morphogenesis of the caudal neural tube and cloacal derivatives from the posterior growth zone.', *Development* **138**(16), 3451–62.

Van Der Maaten, L. & Hinton, G. (2008), 'Visualizing Data using t-SNE', *Journal of Machine Learning Research* **9**, 2579–2605.

van Nimwegen, E. (2006), Scaling laws in the functional content of genomes, *in* 'Power Laws, Scale-Free Networks and Genome Biology', Springer, pp. 236–253.

Verd, B., Clark, E., Wotton, K. R., Janssens, H., Jiménez-Guri, E., Crombach, A. & Jaeger, J. (2018), 'A damped oscillator imposes temporal order on posterior gap gene expression in drosophila', *PLOS Biology* **16**(2), e2003174.

Vetterling, W. T., Flannery, B. P., Press, W. H. & Teuloski, S. A. (1989), *Numerical Recipes in Fortran-The art of scientific computing*, University Press.

Waddington, C. H. (1957), *The strategy of the genes. A discussion of some aspects of theoretical biology. With an appendix by H. Kacser.*, London: George Allen & Unwin, Ltd.

Wallace, E. W. (2010), 'A simplified derivation of the linear noise approximation', *arXiv* .

Watson, J. & Crick, F. (1958), On protein synthesis, *in* 'The Symposia of the Society for Experimental Biology', Vol. 12, pp. 138–163.

Weinstein, N. & Mendoza, L. (2013), 'A network model for the specification of vulval precursor cells and cell fusion control in caenorhabditis elegans', *Frontiers in genetics* **4**, 112.

Weinstein, N., Ortiz-Gutiérrez, E., Muñoz, S., Rosenblueth, D. A., Álvarez-Buylla, E. R. & Mendoza, L. (2015), 'A model of the regulatory network involved in the control of the cell cycle and cell differentiation in the Caenorhabditis elegans vulva', *BMC Bioinformatics* **16**(1), 81.

Wilkinson, D. J. (2011), *Stochastic modelling for systems biology*, CRC press.

Wilson, S. I., Graziano, E., Harland, R., Jessell, T. M. & Edlund, T. (2000), 'An early requirement for FGF signalling in the acquisition of neural cell fate in the chick embryo.', *Current Biology* **10**(8), 421–9.

Wilson, V., Olivera-Martinez, I. & Storey, K. G. (2009), 'Stem cells, signals and vertebrate body axis extension', *Development* **136**.

Wolpert, L., Tickle, C. & Arias, A. M. (2015), *Principles of development*, Oxford University Press, USA.

Yamaguchi, T. P., Takada, S., Yoshikawa, Y., Wu, N. & McMahon, A. P. (1999), 'T (Brachyury) is a direct target of Wnt3a during paraxial mesoderm specification.', *Genes & development* **13**(24), 3185–90.

Zeeman, E. (1976), 'Catastrophe theory', *Scientific American* **234**(4), 65–83.

Zeeman, E. C. (1977), *Catastrophe theory: Selected papers, 1972–1977.*, Addison-Wesley.

Zhou, J. X., Aliyu, M., Aurell, E. & Huang, S. (2012), 'Quasi-potential landscape in complex multi-stable systems', *Journal of the Royal Society Interface* **9**(77), 3539–3553.